International Journal of Audiology 2014; 53: 2–20

informa
healthcare

Discussion Paper

Does the acceptable noise level (ANL) predict hearing-aid use?

Steen Østergaard Olsen* & K. Jonas Brännström†

*Research Laboratory, Department of Otorhinolaryngology, Head and Neck Surgery, Copenhagen University Hospital, Rigshospitalet, Denmark, †Department of Logopedics, Phoniatrics and Audiology, Clinical Sciences, Lund University, Lund, Sweden

Abstract

Objective: It has been suggested that individuals have an inherent acceptance of noise in the presence of speech, and that different acceptance of noise results in different hearing-aid (HA) use. The acceptable noise level (ANL) has been proposed for measurement of this property. It has been claimed that the ANL magnitude can predict hearing-aid use patterns. Many papers have been published reporting on different aspects of ANL, but none have challenged the predictive power of ANL. The purpose of this study was to discuss whether ANL can predict HA use and how more reliable ANL results can be obtained. *Design:* Relevant literature regarding the ANL was found on Medline, Embase, and Google Scholar. Additional information was found as references in the included papers and through personal contacts, for instance when attending audiology conferences. *Study sample:* Forty-five papers published in peer reviewed journals as well as a number of papers from trade journals, posters and oral presentations from audiology conventions. *Conclusions:* An inherent acceptance of noise in the presence of speech may exist, but no method for precise measurement of ANL is available. The ANL model for prediction of HA use has yet to be proven valid.

Key Words: Acceptable noise level; ANL; accuracy; precision; repeatability; prediction; hearing-aid use

The interest in the acceptable noise level (ANL) test has increased among clinicians and researchers alike during the last years. It is a test that is attractive to the clinician as it takes little time to perform and some studies suggest that it can be used to predict future HA use. In two previous studies (Olsen et al, 2012a,b), we have demonstrated that at least the Danish and non-semantic versions of the ANL generate unreliable results that cannot predict HA use.

Nabelek et al (1991) argued that the signal-to-noise ratios (SNR) at which an individual stops attending a conversation might be relevant for prediction of HA use. Consequently they proposed a procedure for measuring the amount of background noise that individuals can tolerate when listening to speech. According to their findings, full-time HA users tolerate more background noise than part-time and non-users of HA when listening to speech. Later Rogers et al (2003) termed the method the acceptable noise level (ANL) test. Several researchers have suggested that the ANL is inherent to an individual (Nabelek et al, 1991, 2006; Rogers et al, 2003; Tampas & Harkrider, 2006) and can be established prior to HA fitting as a possible predictor of hearing-aid use (Nabelek et al, 2004). In a large group of individuals (n = 191), Nabelek et al (2006) found that the ANL measure predicted HA use with 85% accuracy.

The purpose of the present study is to extend the discussion on whether the ANL can be measured accurately using any ANL version and whether ANL can provide reliable results that can predict HA use. Additionally it is discussed how more reliable ANL results can be achieved by improved procedures.

Material and Method

To create a basis for a discussion of the topic: 'Can the ANL be accurately measured and can ANL predict HA use?' we searched for relevant peer reviewed papers on Medline and Embase. The following Boolean search terms were used: 'acceptable AND noise AND level', 'acceptance AND background AND noise' and 'toleration AND background AND noise'. The search was limited to papers published between January 1st, 1991 and December 31st, 2012. Additional papers were found as references in the included papers and through personal contacts. Forty-five peer reviewed relevant papers were identified (Table 1). In order to bring forward further input to the discussion, we searched Google Scholar for information from trade journals and dissertations. In this paper we also refer to posters and oral presentations from audiology conventions.

In the following chapters we discuss various aspects of ANL and finally we propose a conceptual model of how individuals make their ANL judgements.

Discussion

The ANL procedure

Typically ANL is measured by delivering an audio-recorded story to a transducer through one channel of the audiometer. Each individual is asked to use an up-and-down approach to adjust the sound level of the story to the most comfortable level (MCL), by first increasing the

Correspondence: Steen Østergaard Olsen Research Laboratory, Department of Otorhinolaryngology, Head and Neck Surgery, University Hospital, Rigshospitalet, DK-2100 Copenhagen, Denmark. E-mail: steen.oestergaard.olsen@ regionh.dk

(Received 1 February 2013; accepted 26 August 2013)

ISSN 1499-2027 print/ISSN 1708-8186 online © 2014 British Society of Audiology, International Society of Audiology, and Nordic Audiological Society
DOI: 10.3109/14992027.2013.839887

Abbreviations

ADHD	Attention deficit/hyperactivity disorder
ANL	Acceptable noise level
APHAB	Abbreviated profile of hearing-aid benefit
BNL	Background noise level
CR	Coefficient of repeatability
HA	Hearing aid
HTL	Hearing thresholds
ICC	Intraclass correlation coefficients
IOI-HA	International outcome inventory for hearing aids
ISTS	International speech test signal
LDL	Loudness discomfort level
MCID	Minimal clinically important difference
MCL	Most comfortable level
SD	Standard deviation
SNR	Signal-to-noise ratio
WMC	Working memory capacity

loudness until the speech signal becomes too loud, then decreasing it until it becomes too soft, and finally selecting the MCL. When the MCL for speech has been found, a noise signal is delivered to the same transducer through the other channel of the audiometer. The individual is asked to adjust the noise to an acceptable background noise level (BNL). While the speech signal remains at the established MCL the individual is asked to increase the loudness of the noise until it becomes too loud, then to decrease it until the speech becomes very clear, and finally the individual selects the BNL. Nabelek et al (2004) described the BNL as a maximum level of the background noise to which the participant would be willing to accept or 'put up with' without becoming tense and tired while listening to and following the words of the story. The ANL is calculated by subtracting the BNL from the MCL (ANL = MCL − BNL). A high ANL value means that the tested individual accepts little noise compared to the speech level (the difference between MCL and BNL is big) and a low ANL value means that the individual accepts much noise (the difference between MCL and BNL is small).

Goldman (2009) showed that it is mandatory to follow the procedure mentioned above as the testing method used to establish the BNL (ascending or descending noise) has an impact on the ANL outcome.

Calculating the mean of replicated ANL measurements could help obtain an ANL as close as possible to the 'true' ANL (Altman, 1991). In 45 peer-reviewed ANL studies (Table 1), the MCL was established using a single measurement (n = 16), two replications (n = 13), three replications (n = 10). In the last study, a preset fixed speech presentation level was used for all measurements. Five articles did not report the number of MCL replications used. The BNL has been established using a single measurement (n = 6), two replications (n = 14), three replications (n = 20), or four replications (n = 1). Four papers did not report the number of BNL replications used. Notably, seven articles state that one or more replications were excluded before the ANL was calculated (Freyaldenhoven et al, 2005a, 2007; Plyler et al, 2007; Donaldson et al, 2009; Peeters et al, 2009; Rishiq et al, 2012). Such an approach might introduce a bias in the results. Holm & Kastberg (2012) showed improved repeatability by increasing the number of replications to twelve.

Summary

The ANL procedure is well established with regard to most of the included steps, but replicated measurements are—in spite of the known benefits—not always used. Exclusion of some of the measured replications may introduce bias.

Speech materials

ANL has been introduced in several languages: American English (Nabelek et al, 2004), Australian English (Walravens et al, 2012), German (Fredelake et al, 2012), Korean (von Hapsburg & Bahng, 2006), Mandarin (Chen et al, 2011), and Swedish and Danish (Brännström et al, 2012a). Two studies (Brännström et al, 2012a; Ho et al, 2013) have investigated the usability of a speech like signal without semantic content called the international speech test signal (ISTS; Holube et al, 2010). Gordon-Hickey & Moore (2008) also investigated ANL with two speech signals that were unintelligible to the listener (a backward recording of the commonly used American English speech material and a recording of conversational Chinese). Forty of the studies in Table 1 were carried out with English speech material.

When performing the ANL, the individual's comprehension of the story is not tested and according to Goldman (2009) the degree of attention the individual pays to the story does not influence the ANL. As mentioned, unintelligible speech signals or speech-like signals have been used for ANL measurements, but also excerpts from novels recorded as audio books (Brännström et al, 2012a; Walravens et al, 2012) have been used. von Hapsburg & Bahng (2006) recorded a Korean children's story for the ANL measurement. Recordings of non-fictional literature, like descriptions of geography, history or travel have also been used (Nabelek et al, 2004; Plyler et al, 2011; Brännström et al, 2012a). For the German ANL, Fredelake et al (2012) used sentences consisting of five words with a fixed syntactical structure (name-verb-number-adjective-object). Each sentence was grammatically correct but semantically unpredictable.

Chen et al (2011) found no difference between ANLs obtained with Mandarin speech materials which were rated as being 'easy' or 'difficult' in content. Plyler et al (2011) reported that the content of the samples used as speech signal did not significantly affect ANL. von Hapsburg & Bahng (2006) investigated potential language effects on ANL measured in bilinguals and found no differences between ANLs measured with English and Korean speech materials. In a normal-hearing population Brännström et al (2012a) found a trend that ANL was lower when assessed using ISTS as speech than with Danish or Swedish speech signals. In contrast, Gordon-Hickey & Moore (2008) found that ANL increased significantly using reversed or unfamiliar language as speech signal compared to intelligible speech, while Ho et al (2013) found no difference between ANLs measured with ISTS and the listener's first-language (English or Mandarin). Based on these results it cannot be concluded whether the ANL measure is independent of language and of semantic content or not.

Nabelek et al (1991) used a recording of a female speaker as the speech signal, but since then, male speaker recordings have most often been used. Plyler et al (2011) found no significant effect of speaker gender on the ANL, while Gordon-Hickey et al (2012a) found a statistically significant but clinically insignificant speaker gender effect on ANL (< 1 dB). Goldman (2009) showed that listeners accepted more background noise when listening to clear speech (Helfer, 1997) than when listening to speech with normal or fast speech rate. In agreement with this, Brännström et al (2012a) found

Table 1. Forty-five studies published in peer-reviewed journals with information on language version, number of subjects, subjects' hearing status, transducers, and presentation mode. NH = normal hearing, HI = hearing impaired, CI = cochlear implant.

Study	*Language*	*Number of subjects*	*Hearing status*	*Transducer*	*Presentation*
Nabelek et al, 1991	English	75	NH/HI	Earphones	Monaural
Crowley& Nabelek, 1996	English	46	HI	Earphones	Monaural
Rogers et al, 2003	English	50	NH	Loudspeaker	Binaural
Nabelek et al, 2004	English	50	HI	Loudspeaker	Binaural
Freyaldenhoven et al, 2005a	English	40	HI	Loudspeaker	Binaural
Freyaldenhoven et al, 2005b	English	15	NH	Loudspeaker	Binaural
Harkrider & Smith, 2005	English	31	NH	Earphones	Monaural and dichotic
Franklin et al, 2006	English	20	NH	Loudspeaker	Binaural
Freyaldenhoven et al, 2006a	English	30	NH	Loudspeaker	Binaural
Freyaldenhoven et al, 2006b	English	39	HI	Loudspeaker	Binaural
Freyaldenhoven et al, 2006c	English	19	HI	Loudspeaker	Binaural
Harkrider & Tampas, 2006	English	13	NH	Earphones	Diotic
Mueller et al, 2006	English	22	HI	Loudspeaker	Binaural
Nabelek et al, 2006	English	191	HI	Loudspeaker	Binaural
Tampas & Harkrider, 2006	English	21	NH	Earphones	Diotic
von Hapsburg & Bahng, 2006	English, Korean	30	NH	Loudspeaker	Binaural
Freyaldenhoven et al, 2007	English	99	NH/HI	Loudspeaker	Binaural
Gordon-Hickey & Moore, 2007	English	24	NH	Loudspeaker	Binaural
Plyler et al, 2007	English	40	NH/HI	Loudspeaker	Binaural
Freyaldenhoven et al, 2008a	English	191	HI	Loudspeaker	Binaural
Freyaldenhoven et al, 2008b	English	69	HI	Loudspeaker	Binaural
Gordon-Hickey & Moore, 2008	English	30	NH	Loudspeaker	Binaural
Plyler et al, 2008	English	24	NH/CI	Loudspeaker	Binaural
Ahlstrom et al, 2009	English	21	HI	Loudspeaker	Binaural
Donaldson et al, 2009	English	20	CI	Loudspeaker	Binaural
Johnson et al, 2009	English	15	HI	Loudspeaker	Binaural
Peeters et al, 2009	English	18	HI	Loudspeaker	Binaural
Adams et al, 2010	English	24	NH	Loudspeaker	Binaural
Julstrom et al, 2011	English	57	HI	Telecoil	Hearing aid
Kim & Bryan, 2011	English	15	HI	Loudspeaker	Binaural
Moore et al, 2011	English	68	NH	Earphones	Diotic
Plyler et al, 2011	English	43	NH	Loudspeaker	Binaural
Chen et al, 2011	English, Mandarin	31	NH	Loudspeaker	Binaural
Brännström et al, 2012a	Danish, Swedish, ISTS	80	NH	Earphones	Monaural
Brännström et al, 2012b	Swedish	21	NH	Earphones	Monaural
Franklin et al, 2012	English	25	NH	Loudspeaker	Binaural
Fredelake et al, 2012	German	21	NH/HI	Earphones	Diotic
Gordon-Hickey et al, 2012a	English	30	NH	Loudspeaker	Binaural
Gordon-Hickey et al, 2012b	English	25	NH	Earphones	Diotic
Nichols & Gordon-Hickey, 2012	English	70	NH	Loudspeaker	Binaural
Olsen et al, 2012a	Danish, ISTS	40	NH	Earphones	Monaural
Olsen et al, 2012b	Danish, ISTS	63	HI	Earphones	Monaural
Rishiq et al, 2012	English	19	NH	Earphones	Monaural
Recker & Edwards, 2013	English	20	NH/HI	Loudspeaker	Binaural
Ho et al, 2013	English, Mandarin, ISTS	60	NH	Loudspeaker	Binaural

that Swedish individuals accepted more noise while listening to a Swedish text presented at 3 syllables per second than Danish individuals listening to a Danish text presented at 4 syllables per second. However, Ho et al (2013) did not observe a similar trend between American English (4.5 syllables per second) and Mandarin (3.6 syllables per second).

Summary

Most ANL studies have been carried out using English speech materials. Speaker gender does not affect the ANL, but the speech rate and the semantic content may have influence on the ANL outcome.

Noise materials

Among the studies listed in Table 1, forty-one studies used the conventional multi-talker babble for the BNL measurement, while three studies (Mueller et al, 2006; Peeters et al, 2009; Fredelake et al, 2012) used non-modulated speech-shaped noise. These three studies investigated the effect of a noise reduction algorithm which reduces sound with little fluctuation (labeled as noise). This might be the reason for selecting this type of noise signal. Of the 41 studies using multi-talker babble as noise signal, the effect of other noise types were also investigated in seven studies (Nabelek et al, 1991; Freyaldenhoven et al, 2006a; Gordon-Hickey & Moore, 2007; Brännström et al, 2012a; Gordon-Hickey et al, 2012a; Olsen et al, 2012a,b). The additional signals used as competing noise were: Non-modulated or modulated speech-shaped noise, traffic noise, music, pneumatic drill, one person talking (forward and backward recording), and four-talker babble.

Nabelek et al (1991) found no effect of noise type on ANL except for music, and Crowley& Nabelek (1996) found that the difference between ANLs measured with 12-speaker babble and steady state speech-shaped noise was less than 1 dB. Other studies have reported that noise type influences ANLs. For instance, mean ANLs obtained by Freyaldenhoven et al (2006a) using speech-babble noise were approximately 2 dB lower than mean ANLs obtained using speech-spectrum noise. Gordon-Hickey & Moore (2007) compared ANL with the conventional 12-talker babble with six music samples as noise signals. The mean ANL found with music was 3.7 dB lower than the ANL found with speech babble, but ANL for music was not related to music preference. Gordon-Hickey et al (2012a) found only a small effect on the mean ANL (about 1 dB) when comparing a one-talker masker with a multi-talker masker. Gordon-Hickey et al (2012a) also found a 0.86 dB lower mean ANL with backward recorded speech as compared to intelligible speech. According to the authors, intelligible one-talker background acts primarily as an informational masker yielding higher ANLs.

In two groups of normally-hearing subjects, Brännström et al (2012a) found significantly lower mean ANLs measured with modulated speech-weighted noise than with speech-babble. The difference was about 4 dB. In a study of hearing-impaired subjects (Olsen et al, 2012b), executed in one of the facilities used by Brännström et al (2012a), no difference between ANLs measured with the same two stimuli was found. Referring to the ‘glimpse effect’ (Cooke, 2006), Phatak & Grant (2012) found that hearing-impaired listeners achieved as much modulation benefit as normal-hearing listeners for slower masker modulation rates (2, 4, and 12 Hz), but showed a reduced benefit for the fast masker modulation rate of 32 Hz. If the ‘glimpse effect’ in contrast to what was suggested by Gordon-Hickey et al (2012a) also has an influence on ANL it may explain both the ANL differences found in normal-hearing subjects and that the ANLs are the same in hearing-impaired subjects when measured with speech-babble and with modulated speech shaped noise.

Summary

In most studies speech babble has been used as noise signal, which seems to be warranted since most studies comparing different noise materials for ANL measurement shows that the noise material has an influence on the ANL outcome. ANL values may depend on whether the masking is informational or energetic, and whether the masking is steady state or fluctuating.

Presentation mode

Nabelek et al (1991) presented the speech and noise signals monaurally through earphones, but in most studies (31 of 45 articles we have reviewed, Table 1) the signals were presented in the sound field through loudspeaker(s). However, insertion, supra-aural, and circumaural earphones have been used (e.g. Harkrider & Smith, 2005; Tampas & Harkrider, 2006; Harkrider & Tampas, 2006; Moore et al, 2011; Brännström et al, 2012a). In most of the peer reviewed studies, the test was conducted binaurally (31 studies), but monaural (eight studies), diotic (five studies), and dichotic (one study) presentations have also been used. Freyaldenhoven et al (2006b) demonstrated that ANLs measured through monaural amplification are similar to those obtained using binaural amplification. No study has directly investigated differences between monaural, diotic, and binaural ANLs. Dichotic presentation yields the largest ANL range, − 18 to 40 dB (Harkrider & Smith, 2005). Also, different versions of ANL seem to provide differences in absolute ANL values (Nabelek et al, 1991; Rogers et al, 2003; Franklin et al, 2006; Freyaldenhoven et al, 2007; Harkrider & Tampas, 2006; Tampas & Harkrider, 2006; von Hapsburg & Bahng, 2006; Gordon-Hickey & Moore, 2007; Plyler et al, 2008). Presentation mode yields differences in ANLs, but we argue that it should not affect the relationship between ANL and HA use. Any association should present itself irrespective of absolute ANLs. However, if ANL data are collected differently than done by Nabelek et al, (2006) who presented signals through a front loudspeaker, these authors recommend developing new baseline data.

Summary

Several transducer types have been used for the presentation of ANL signals. If transducers other than a front loudspeaker are used, normative data for the specific transducer should be used.

Response modes

Hand signs or hand-held buttons allowing the test subject to signal the experimenter to adjust the levels of speech and noise have been used in most studies listed in Table 1 (n = 32). In other studies individuals have adjusted the levels themselves using attenuators, audiometer controls, or computerized potentiometers. Most notably, in the seminal study by Nabelek et al (1991) the individuals adjusted the levels themselves and had visual feedback from audiometer dials. Although each individual was instructed to disregard the markings on the audiometer dials, the visibility of the settings represents an important bias. A hand sign that the experimenter must interpret also implies bias on the ANL results. In our opinion procedures in which individuals adjust the dials without visual feedback with a final ready sign from the individual (Brännström et al, 2012a) or using a personal computer equipped with a custom graphical user interface (Recker & Edwards, 2013) have the least bias and should be preferred.

Summary

In most studies test subjects signalled the experimenter to increase or decrease the levels of the test signals. Direct adjustments of the levels by the test subjects should be preferred.

Instructions

As with any psycho-physical method, the instructions given are very important for the outcome. Rogers et al (2003) were the first to publish written and verbal ANL instructions. Freyaldenhoven et al (2005a) published the instruction set that is now most often used and which is included in the accompanying leaflet to the ANL CD (Acceptable Noise Level Test, Cosmos Distributing Inc).

The instructions read:

For the MCL: ‘You will listen to a story through a loudspeaker. After a few moments, select the loudness of the story that is most comfortable for you, as if listening to a radio. Hand-held buttons will allow you to make adjustments. First, turn the loudness up until it is too loud and then down until it is too soft. Finally, select the loudness level that is most comfortable for you. ’

For the BNL: ‘You will listen to the same story with background noise of several people talking at the same time. After you have listened to this for a few moments select the level of background noise that is the MOST you would be willing to accept or 'put up with' without becoming tense or tired while following the story. First, turn the noise up until it is too loud and then down until the story becomes very clear. Finally, adjust the noise (up and down) to the MAXIMUM noise level that you would be willing to 'put up with' for a long time while following the story.'

The instruction set is often modified to fit individual needs. For example this is the case when listeners do not use hand-held buttons to respond or when the noise is not babble. Such modifications may have an influence on the ANL outcome and should therefore be limited to the absolutely necessary.

Of the papers listed in Table 1, thirty-four papers include the instruction text or give a reference to a specific paper in which the text can be found. From reading the papers, it is not always clear if instructions were given orally, in writing or both. In eleven papers the content of the instructions was not revealed. This may cause problems, for instance when ANL results are compared across studies and it is not possible to consider the influence from varying instruction sets.

Translation and interpretation of the instructions could have a large effect on the ANL outcome. Administering the ANL with translated instructions should be done with caution (Ho et al, 2013). Ho et al (2013) and Brännström et al (2012a) found that extrinsic factors, such as examiner attitude, and the listener's cultural background, might cause ANL differences between groups of individuals.

Summary

Most studies have used the instruction set by Freyaldenhoven et al (2005a), which should be recommended in order to ensure that ANL measurements can be compared to the widest possible extent. Translations should be faithful not only to the wording but also to the characteristic of the ANL measure.

Factors related to ANL

Several studies suggest that ANL is inherent to the individual (Nabelek et al, 1991, 2006; Rogers et al, 2003; Tampas & Harkrider, 2006), and this would, for example, mean that ANL should remain constant regardless of whether the right or the left ear is measured. However, Freyaldenhoven et al (2006b) reported differences of up to 6 dB in ANL between ears in four individuals. Rather than taking this report as evidence that individuals might have different ANL on the right and the left ear as suggested by Freyaldenhoven et al (2006b), we believe that these discrepancies most probably occur due to the poor precision of ANL measurements (see below).

Measurements of the ANL show very large variations across individuals. Some factors which do not affect ANL include: age (Nabelek et al, 1991; Moore et al, 2011), gender (Rogers et al, 2003), the loudness discomfort level (LDL) in individuals with normal hearing (Franklin et al, 2012) and speech scores in noise (Nabelek et al, 2004). Activity levels in the medial olivocochlear bundle system have no influence on ANL outcome (Harkrider & Smith, 2005). Nabelek et al (1991, 2006) reported that degree of hearing loss does not affect ANLs, while Walravens et al (2012) showed a weak but statistically significant association between ANL and hearing loss in a study with 290 individuals.

Franklin et al (2006) and Freyaldenhoven et al (2007) demonstrated that ANL increases with increasing MCL. Brännström et al (2012b) investigated a possible association between working memory capacity (WMC) and ANL. A strong association was found between BNL and WMC while a much weaker but still statistically significant association was found between ANL and WMC. Freyaldenhoven et al (2005a) found that ANL decreased when female college students with attention deficit/hyperactivity disorder (ADHD) were medicated with stimulant drugs. The authors suggested that this was caused by changes in auditory processing resulting from central processes like suppression of cortical activity or enhanced inhibitory processes. Harkrider & Tampas (2006) and Tampas & Harkrider (2006) found differences in the auditory brainstem response, auditory middle latency response, and cortical, auditory late latency response between a group of females with high ANLs and a group of females with low ANLs. According to these authors, central cognitive processes govern the acceptance of background noise when listening to speech.

Nichols & Gordon-Hickey (2012) investigated the influence of two specific personality characteristics on ANLs: locus of control and self-control. Individuals who view events as the result of their own actions have an internal locus of control, while individuals who view events as the result of luck or fate have an external locus of control. Nichols & Gordon-Hickey (2012) found no significant association between locus of control and ANLs. However, greater self-control, which refers to the ability to control one's own thoughts, emotions, impulses, and performance, was found to be associated with lower ANLs (Nichols & Gordon-Hickey, 2012). Still, self-control only accounted for 8% of the ANL variance. Thus, psychological factors influence the ANL.

To prove that ANL is inherent to the individual it would be necessary to carry out a prospective ANL study showing that the individual ANL remains the same through life, independent of changing life conditions.

Summary

ANL is not influenced by age, gender, LDL, speech in noise scores, and activity levels in the medial olivocochlear bundle system. It is uncertain whether hearing sensitivity has influence on the ANL. ANL is associated with the WMC and is influenced by the intake of stimulant drugs, by psychological factors, and by the speech presentation level. Central cognitive processes govern ANL. Whether ANL is an inherent property has not yet been proven.

Accuracy and precision of ANL measurements

ANL is a psychoacoustic test which is limited by two central features encountered in all psychoacoustic testing: accuracy and precision. Measurement accuracy is the closeness of agreement between a measured quantity value and a true quantity value (ISO/IEC, 2007). The ANL is the only method for measurement of the acceptance of noise in the presence of speech. It is therefore difficult to estimate the accuracy of ANL measurements.

Measurement precision (or repeatability) is the closeness of agreement between measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions (ISO/IEC, 2007). The results of a hearing test from a certain individual should be similar when repeated by the same or by other testers. In some earlier studies on repeatability of pure-tone audiometry (Jerlvall & Arlinger, 1986; Beahan et al, 2012; Kam et al, 2012) precision was reported as the percentage of agreement between the thresholds obtained at two measurements and/or as the SD of the differences between measurements. The coefficient of repeatability (CR; Bland & Altman, 1986), which is $1.96 \times$ SD of the differences between the repeated measurements, is also a measure of precision. This means that for a normal distribution, 95% of the absolute differences will be less than the CR. In this manner the CR estimates the agreement between repeated individual measurements, which is needed to evaluate the precision of ANL. The minimal clinically important difference (MCID) is the smallest relevant or important difference between measures (Beaton et al, 2002) and it must be taken into consideration when evaluating the precision of a method. The MCID is not a value that can be calculated; rather it has to be estimated for each type of measurement based on the property that is going to be measured. HA directional microphones and noise reduction algorithms reduce ANL by about 3–4 dB (Freyaldenhoven et al, 2005b; Mueller et al, 2006; Kim & Bryan, 2011). Freyaldenhoven et al (2006a) stated that a 2-dB difference in ANL may not be clinically relevant, Ho et al (2013) used 3 dB as the clinically significant difference, and according to Gordon-Hickey et al (2012a) a 0.66-dB difference in ANL may not be clinically significant for measurement of ANL. If the MCID based on these data is selected as 3 dB, then to deem the ANL repeatable, the CR should be less than 3 dB.

According to Nabelek et al (2006) individuals can be separated into three ANL categories: (1) Individuals with ANL < 7 dB will become successful users (HAs used whenever needed), (2) individuals with ANL > 13 dB will become unsuccessful users (HAs used occasionally / HAs no longer used), (3) for individuals with ANL between 7–13 dB, HA use cannot be predicted. An ANL CR at 3 dB is much less than the 6 dB that separates ANLs predicting successful and unsuccessful hearing-aid use and hence it could be avoided that measurements from one occasion would predict success and a repeated measurement would predict lack of success just because of poor precision.

Table 2 shows the CRs reported from four studies (Olsen et al, 2012a,b; Holm & Kastberg, 2012; Walravens et al, 2012) and from one study (Gordon-Hickey et al, 2012a) that did not report the CR, but in which case we in an earlier paper (Olsen et al, 2013) were able to calculate the CR based on the data given in the original paper. Along with the CRs the maximum differences between repeated measures from the papers are shown. Two other papers (Freyaldenhoven et al, 2006a; Nabelek et al, 2004) did not report CRs or maximum differences, but in a later paper Nabelek et al (2007) reported that in these two mentioned studies the maximum differences shown in Table 2 were found. All but one study (Gordon-Hickey et al, 2012a) reported data obtained with the same test subjects and testers at different test sessions. Gordon-Hickey et al (2012a) compared measurements at three test sessions on the same day with the same test subjects but with three different testers. It is clear that most studies in Table 2 report magnitudes of CRs and/or differences between repeated measures, which do not allow for reliable ANLs. One study (Nabelek et al, 2004) reports remarkably low test-retest differences compared to the other studies. A possible reason for the small differences might be that the individuals had previously participated in ANL studies, while the individuals in the other studies were completing the ANL for the first time.

Summary

The accuracy of ANL measurements is unknown, while it has been demonstrated that the precision of ANL measurements is too poor to yield repeatability lower than 3 dB.

Prediction of future HA use

It would be valuable to identify a priori factors influencing HA benefit, use time, and satisfaction. Knudsen et al (2010) examined the literature covering studies addressing four variables: help seeking, hearing-aid uptake, hearing-aid use, and satisfaction with hearing aids. Knudsen et al (2010) found that among 31 quoted factors, only one—self-reported hearing disability—positively affected all four outcome variables.

Nabelek et al (2006) estimated that ANL alone can predict HA use with 85% accuracy. However in a letter to the editor, Nabelek et al (2007) wrote that the ANL accuracy in predicting hearing aid success was 82%, while the precision for predicting lack of success was 83.6%, and that about 25% of listeners therefore would be misclassified with the ANL procedure. Further, Nabelek et al (2006) stated that a statistical model tend to fit the data upon which

Table 2. The coefficients of repeatability (CR) from five studies are shown, along with the maximum difference between the first and the last measured ANL from six studies. The CR and the maximum differences between measurements from the study by Gordon-Hickey et al (2012b) was calculated by us using data from tables and figures in the original study. The maximum differences between measurements in the studies by Freyaldenhoven et al (2006a) and Nabelek et al (2004) were reported by Nabelek (2007). The repeatability is shown for six studies with the same tester (intra tester) and one study with different testers (inter tester).

				Sessions on the same day			*Sessions on separate days*	
	Study	*Number of subjects*		*Normal hearing*	*Mixed*	*Impaired hearing*	*Normal hearing*	*Impaired hearing*
Intra tester	Freyaldenhoven et al (2006a)	30	CR (dB)				–	
			Maximum test-retest difference (dB)				14.3	
	Nabelek et al (2004)	50	CR (dB)					–
			Maximum test-retest difference (dB)					4.0
	Olsen et al (2012a)	39	CR (dB)	6.0–8.9			8.8–10.2	
			Maximum test-retest difference (dB)	12.0			15.3	
	Olsen et al (2012b)	63	CR (dB)			6.5–8.6		7.1–8.8
	Holm & Kastberg (2012)	32	CR (dB)	7.6				
			Maximum test-retest difference (dB)	14.0				
	Walravens et al (2012)	290	CR (dB)		8.5			
			Max test retest diff (dB)		20			
Inter tester	Gordon-Hickey et al (2012a)	25	CR (dB)	4.7–7.6				
			Max test retest diff (dB)	8.7				

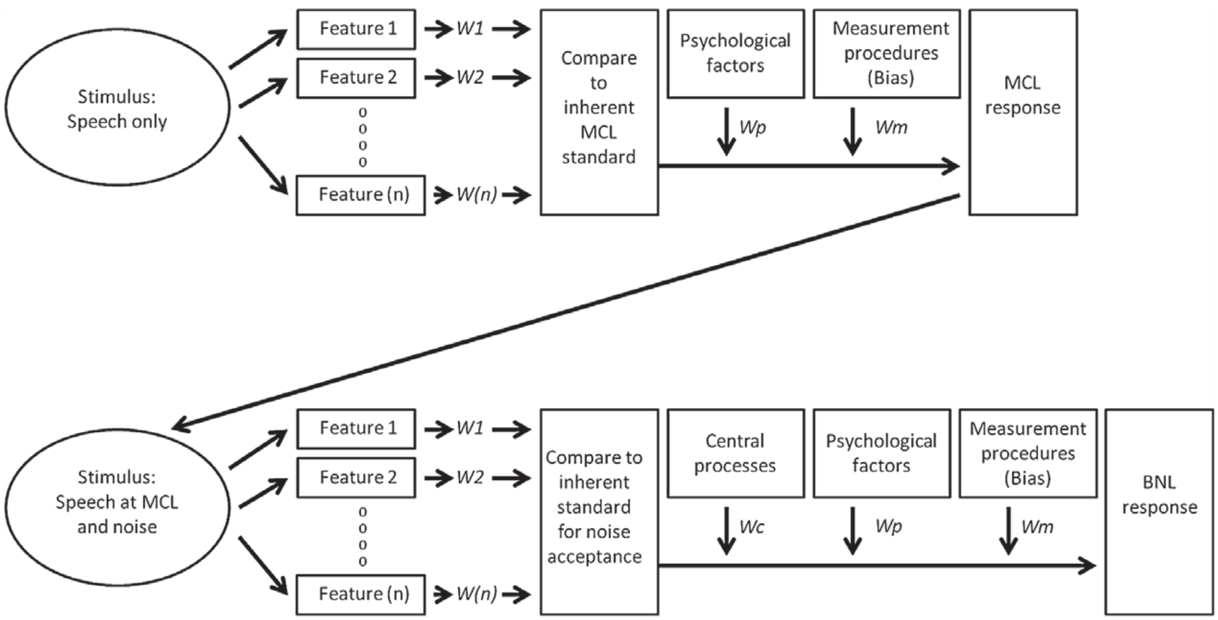

Figure 1. A revised model of how listeners make their ANL judgements based on a model by Wu et al (2013, personal communication). The first stage is to establish the MCL by comparing the acoustic features of the speech stimuli to the listeners' own inherent standard. The MCL is influenced by psychological factors and by measurement procedures. All factors have individual weights (*W*). Since the same individual might label a whole range of sound levels as the MCL and because ANL increases with increasing MCL, the sound level of the selected MCL may be one of the acoustic features affecting the ANL. WMC and central auditory processes play a role for the BNL as well as psychological factors. Finally the measurement procedures introduce bias. All factors have individual weights (*W*).

they are developed better than on independent data. The ANL model has never been tested on independent data, which would be needed to show its predictive power (Schwartz & Cox, 2012). It would be interesting to investigate whether ANL can predict HA use in a naïve group of hearing-impaired individuals before they are fitted with their first HA.

As stated before, the ANL model predicts that individuals with ANL < 7 dB will become successful users (HAs used whenever needed), and that individuals with ANL > 13 dB will become unsuccessful users (HAs used occasionally / HAs no longer used), but that HA use not can be predicted for individuals with ANL between 7–13 dB. In the paper, HA use was based on hearing-impaired individuals' responses to a questionnaire which reads: 'How do you use your HAs? (check 1, 2, or 3)'. 1. I wear my HAs whenever I need them (Approximately how many hours?); 2. I only wear my HAs occasionally (Approximately how many hours?/Why? Briefly describe the situations); and 3. I do not wear my HAs (Why do you not wear them?). We argue, that the statement 'I wear my HAs whenever I need them' defining the group of successful users is rather vague and that the unsuccessful group collapsed of groups 2 and 3 is too heterogeneous. A measure that tells us that a hearing-impaired individual in the future will use their HA 'whenever needed' is not very informative. Before further discussing if ANL accurately predicts future HA use, meaningful HA-use categories should be defined.

In a retrospective study, Schwartz & Cox (2012) questioned the predictive power of ANL, since it could not predict HA-use as measured by five standardized hearing outcome measures (e.g. the abbreviated profile of hearing aid benefit (APHAB); Cox & Alexander, 1995). A possible relationship between APHAB and ANL was also investigated by Freyaldenhoven et al (2008a), who, in agreement with Schwartz & Cox (2012), found no relationship between the two measures. Taylor (2008) found that ANL could predict scores on the international outcome inventory for hearing aids (IOI-HA; Cox et al, 2000), a finding that Olsen et al (2012b) could not replicate. Hartley et al (2010) concluded that since ANLs do not appear to be related to reported hours of usage, they may not be useful in predicting HA utilization in potential users. Finally, it is noteworthy that a recent study (Walravens et al, 2012), in contrast to Nabelek et al (2006), found that unsuccessful HA users have statistically significant *lower* ANL values than successful users ($P < 0.05$).

Summary

The ANL categories are not very informative, and 25% of test subjects would be misclassified by ANL. The predictive power of ANL has never been tested on independent data.

A conceptual ANL model

Numerous studies on ANL have been published, but only recently a study by Wu et al (2013, personal communication) has provided a conceptual model of how individuals make their ANL judgements. The model suggests that individuals, when performing the ANL, use several acoustic features of the stimuli (e.g. speech intelligibility and loudness) which they compare to their inherent standard of noise acceptability when listening to speech. Furthermore, psychological factors (e.g. self-control) influence ANLs. The individuals assign different weights to both the acoustic features and the psychological factors depending on their individual preferences. This might be the explanation for the large ANL variability across individuals. The model suggests that this complex behaviour could explain why the ANL can predict HA use. However, it also suggests that the repeatability will be poor since the weights each individual applies may change over time. Unfortunately, this model does not recognize that there may be a substantial amount of variability induced by the method itself. Therefore, we revised the model as seen in Figure 1.

The model has two stages: (1) Establishment of the MCL, and (2) establishment of the BNL. Each individual finds the MCL after comparing the acoustic features of the speech stimuli to an inherent standard, which is also influenced by psychological factors. The MCL is also influenced by measurement procedures and all the factors have individual weights (*W*). The MCL is not a threshold but rather a range of levels for each individual (Punch et al, 2004). Franklin et al (2006) and Freyaldenhoven et al (2007) found that ANL increases (becomes poorer) with increasing MCL, and therefore the choice of MCL will be one of the acoustic features that affect the ANL. As for the MCL, psychological factors play a role for the BNL as well as WMC and central auditory processes (Freyaldenhoven et al, 2005a; Harkrider & Tampas, 2006; Tampas & Harkrider, 2006; Brännström et al, 2012b). Finally, the measurement procedures introduce bias. All factors have individual weights (*W*). Generally, the model is similar to the one proposed by Wu et al (2013, personal communication) in describing the complex behaviour, but additionally it suggests that the nature of the MCL (as a range of levels) adds to the variability of the ANL measure. The two models yield identical ANL results, but the latter demonstrates why ANL due to large measurement errors may not be related to HA use.

SUMMARY

A conceptual ANL model including the effect of measurement procedures, psychological factors, WMC, and central auditory processes has been introduced.

Conclusions and future research directions

The ANL procedure is well established, but the precision of the existing ANL method is poor, and it has not been demonstrated that ANL is an inherent property. Future research should focus on improving the ANL method in order to improve the repeatability. Such research should focus on the number of replicated measures, the role of instructions, and bias of response modes. The conceptual ANL model revised for the present paper may help to understand how for instance the ANL procedures may affect the precision of the method.

Most ANL studies have been carried out using the English ANL speech babble noise presented through a front loudspeaker (Nabelek et al, 2006) and the instruction set by Freyaldenhoven et al (2005a). When deviating from this setup, results cannot be compared to the baseline data from Nabelek et al (2006). Normative data for other setups should be developed. Great care should be taken in adopting ANL to other languages and cultures.

Since the presently used ANL categories do not carry much information and 25% of test subjects would be misclassified by ANL, new categories should be developed and the predictive power of ANL should be tested on independent data using the new categories.

Acknowledgements

The authors thank Ariane Laplante-Lévesque and Claus Elberling for valuable comments on a draft of this paper.

Declaration of interest: No outside funding or grants in support of this work were received. The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Adams E.M., Gordon-Hickey S., Moore R.E. & Morlas H. 2010. Effects of reverberation on acceptable noise level measurements in younger and older adults. *Int J Audiol*, 49, 832–838.

Ahlstrom J.B., Horwitz A.R. & Dubno J.R. 2009. Spatial benefit of bilateral hearing aids. *Ear Hear*, 30, 203–218.

Altman D.G. 1991. *Practical Statistics for Medical Research*. London: Chapman & Hall.

Beahan N., Kei J., Driscoll C., Charles B. & Khan A. 2012. High-frequency pure-tone audiometry in children: A test-retest reliability study relative to ototoxic criteria. *Ear Hear*, 33, 104–111.

Beaton D.E., Boers M. & Wells G.A. 2002. Many faces of the minimal clinically important difference (MCID): A literature review and directions for future research. *Curr Opin Rheumatol*, 14, 109–114.

Bland J.M. & Altman, D.G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 327, 307–310.

Brännström K.J., Lantz J., Nielsen L.H. & Olsen S.Ø. 2012a. Acceptable noise level with Danish, Swedish, and non-semantic speech materials. *Int J Audiol*, 51, 146–156.

Brännström K.J., Zunic E., Borovac A. & Ibertsson T. 2012b. Acceptance of background noise, working memory capacity, and auditory evoked potentials in subjects with normal hearing. *J Am Acad Audiol*, 23, 542–552.

Chen J., Zhang H., Plyler P.N., Cao W. & Chen J. 2011. Development and evaluation of the Mandarin speech signal content on the acceptable noise level test in listeners with normal hearing in mainland China. *Int J Audiol*, 50, 354–360.

Cooke M. 2006. A glimpsing model of speech perception in noise. *J Acoust Soc Am*, 119, 1562–1573.

Cox R., Hyde M., Gatehouse S., Noble W., Dillon H. et al. 2000. Optimal outcome measures, research priorities, and international cooperation. *Ear Hear*, 21, 106S–115S.

Cox R.M. & Alexander G.C. 1995. The abbreviated profile of hearing aid benefit. *Ear Hear*, 16, 176–186.

Crowley H.J. & Nabelek I.V. 1996. Estimation of client-assessed hearing aid performance based upon unaided variables. *J Speech Hear Res*, 39, 19–27.

Donaldson G.S., Chisolm T.H., Blasco G.P., Shinnick L.J., Ketter K.J. et al. 2009. BKB-SIN and ANL predict perceived communication ability in cochlear implant users. *Ear Hear*, 30, 401–410.

Franklin C.A., Thelin J.W., Nabelek A.K. & Burchfield S.B. 2006. The effect of speech presentation level on acceptance of background noise in listeners with normal hearing. *J Am Acad Audiol*, 17, 141–146.

Franklin C.A., White L.J. & Franklin T.C. 2012. Relationship between loudness tolerance and the acceptance of background noise for young adults with normal hearing. *Percept Mot Skills*, 114, 717–722.

Fredelake S., Holube I., Schlueter A. & Hansen M. 2012. Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms. *Int J Audiol*, 51, 299–308.

Freyaldenhoven M.C., Nabelek A.K., Burchfield S.B. & Thelin J.W. 2005b. Acceptable noise level as a measure of directional benefit. *J Am Acad Audiol*, 16, 228–236.

Freyaldenhoven M.C., Plyler P.N., Thelin J.W. & Burchfield S.B. 2006b. Acceptance of noise with monaural and binaural amplification. *J Am Acad Audiol*, 17, 659–666.

Freyaldenhoven M.C., Plyler P.N., Thelin J.W. & Hedrick M.S. 2007. The effects of speech presentation level on acceptance of noise in listeners with normal and impaired hearing. *J Speech Hear Res*, 50, 878–885.

Freyaldenhoven M.C., Plyler P.N., Thelin J.W. & Muenchen R.A. 2008b. Acceptance of noise growth patterns in hearing aid users. *J Speech Hear Res*, 51, 126–135.

Freyaldenhoven M.C., Plyler P.N., Thelin J.W., Nabelek A.K. & Burchfield S.B. 2006c. The effects of venting and low-frequency gain compensation on performance in noise with directional hearing instruments. *J Am Acad Audiol*, 17, 168–178.

Freyaldenhoven M.C., Smiley D.F., Muenchen R.A. & Konrad T.N. 2006a. Acceptable noise level: Reliability measures and comparison to preference for background sounds. *J Am Acad Audiol*, 17, 640–648.

Freyaldenhoven M.C., Nabelek A.K. & Tampas J.W. 2008a. Relationship between acceptable noise level and the abbreviated profile of hearing aid benefit. *J Speech Lang Hear Res*, 51, 136–146.

Freyaldenhoven M.C., Thelin J.W., Plyler P.N., Nabelek A.K. & Burchfield S.B. 2005a. Effect of stimulant medication on the acceptance of background noise in individuals with attention deficit/hyperactivity disorder. *J Am Acad Audiol*, 16, 677–686.

Goldman J.J. 2009. The effects of testing method, alternate types of target stimuli and attention on Acceptable Noise Level (ANL) scores in normal hearing listeners. Doctoral dissertation, Communication Sciences and Disorders, James Madison University.

Gordon-Hickey S., Moore R.E. & Estis J.M. 2012a. The impact of listening condition on background noise acceptance for young adults with normal hearing. *J Speech Lang Hear Res*, 55, 1356–1372.

Gordon-Hickey S. & Moore R.E. 2007. Influence of music and music preference on acceptable noise levels in listeners with normal hearing. *J Am Acad Audiol*, 18, 417–427.

Gordon-Hickey S. & Moore R.E. 2008. Acceptance of noise with intelligible, reversed, and unfamiliar primary discourse. *Am J Audiol*, 17, 129–135.

Gordon-Hickey S., Adams E., Moore R., Gaal A. & Berry K. 2012b. Intertester reliability of the acceptable noise level. *J Am Acad Audiol*, 23, 534–541.

Harkrider A.W. & Smith S.B. 2005. Acceptable noise level, phoneme recognition in noise, and measures of auditory efferent activity. *J Am Acad Audiol*, 16, 530–545.

Harkrider A.W. & Tampas J.W. 2006. Differences in responses from the cochleae and central nervous systems of females with low versus high acceptable noise levels. *J Am Acad Audiol*, 17, 667–676.

Hartley D., Wanrooy E., Hickson L. & Mitchell P. 2010. Development of an Australian version of the Acceptable Noise Level (ANL) Test. Audiology Australia, XIX National Conference, 16–19th of May, 2010, Sydney, Australia.

Helfer K.S. Auditory and auditory-visual perception of clear and conversational speech. 1997. *J Speech Hear Res*, 40, 432–443.

Ho H.C., Wu Y.H., Hsiao S.H., Stangl E., Lentz E.J. et al. 2013. The equivalence of acceptable noise level (ANL) with English, Mandarin, and non-semantic speech: A study across the U.S. and Taiwan. *Int J Audiol*, 52, 83–91.

Holm L. & Kastberg T. 2012. Stabilitet för Acceptable Noise Level (ANL) hos normalhörande vuxna personer vid upprepade mätningar inom samma testsession och dess relation till arbetsminneskapacitet. Vetenskapligt arbete, Avdelningen för logopedi, foniatri och audiologi Institutionen för kliniska vetenskaper, Medicinska Fakulteten, Lunds Universitet, Lund. [In Swedish]

Holube I., Fredelake S., Vlaming M. & Kollmeier B. 2010. Development and analysis of an International Speech Test Signal (ISTS). *Int J Audiol*, 49, 891–903.

ISO/IEC Guide 99:2007. 2007. International vocabulary of metrology. Basic and general concepts and associated terms (VIM).

Jerlvall L. & Arlinger S. 1986. A comparison of 2-dB and 5-dB step size in pure-tone audiometry. *Scand Audiol*, 15, 51–56.

Johnson E., Ricketts T. & Hornsby B. 2009. The effect of extending high-frequency bandwidth on the Acceptable Noise Level (ANL) of hearing-impaired listeners. *Int J Audiol*, 48, 353–362.

Julstrom S., Kozma-Spytek L. & Isabelle S. 2011.Telecoil-mode hearing aid compatibility performance requirements for wireless and cordless handsets: Magnetic signal-to-noise. *J Am Acad Audiol*, 22, 528–541.

Kam A.C., Sung J.K., Lee T., Wong T.K. & van Hasselt A. 2012. Clinical evaluation of a computerized self-administered hearing test. *Int J Audiol*, 51, 606–610.

Kim J.S. & Bryan M.F. 2011. The effects of asymmetric directional microphone fittings on acceptance of background noise. *Int J Audiol*, 50, 290–296.

Knudsen L.V., Öberg M., Nielsen C., Naylor G. & Kramer S.E. 2010. Factors influencing help seeking, hearing aid uptake, hearing aid use, and satisfaction with hearing aids: A review of the literature. *Trends Ampl*, 14, 127–154.

Moore R., Gordon-Hickey S. & Jones A. 2011. Most comfortable listening levels, background noise levels, and acceptable noise levels for children and adults with normal hearing. *J Am Acad Audiol*, 22, 286–293.

Mueller H.G., Weber J. & Hornsby B.W. 2006. The effects of digital noise reduction on the acceptance of background noise. *Trends Ampl*, 10, 83–93.

Nabelek A.K., Freyaldenhoven M.C., Tampas J.W., Burchfield S.B. & Muenchen R.A. 2006. Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, 17, 626–639.

Nabelek A.K., Tampas J.W. & Burchfield S.B. 2004. Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. *J Speech Lang Hear Res*, 47, 1001–1011.

Nabelek A.K., Tampas J.W. & Freyaldenhoven M.C. 2007. Further questions about the acceptable noise level test: A response to Dr. Hamill. *J Am Acad Audiol*, 18, 185–187.

Nabelek A.K., Tucker F.M. & Letowski T.R. 1991. Toleration of background noises: Relationship with patterns of hearing aid use by elderly persons. *J Speech Hear Res*, 34, 679–685.

Nichols A.C. & Gordon-Hickey S. 2012. The relationship of locus of control, self-control, and acceptable noise levels for young listeners with normal hearing. *Int J Audiol*, 51, 353–359.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2012a. The Acceptable Noise Level: Repeatability with Danish and non-semantic speech materials for adults with normal hearing. *Int J Audiol*, 51, 557–563.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2012b. Acceptable Noise Level (ANL) with Danish and non-semantic speech materials in adult hearing aid users. *Int J Audiol*, 51, 678–688.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2013. Intertester reliability of the acceptable noise level. *J Am Acad Audiol*, 24, 241–243.

Peeters H., Kuk F., Lau C.C. & Keenan D. 2009. Subjective and objective evaluation of noise management algorithms. *J Am Acad Audiol*, 20, 89–98.

Phatak S.A. & Grant K.W. 2012. Phoneme recognition in modulated maskers by normal-hearing and aided hearing-impaired listeners. *J Acoust Soc Am*, 132, 1646–1654.

Plyler P.N., Alworth L.N., Rossini T.P. & Mapes K.E. 2011. Effects of speech signal content and speaker gender on acceptance of noise in listeners with normal hearing. *Int J Audiol*, 50, 243–248.

Plyler P.N., Bahng J. & Von Hapsburg D. 2008. The acceptance of background noise in adult cochlear implant users. *J Speech Hear Res*, 51, 502–515.

Plyler P.N., Madix S.G., Thelin J.W. & Johnston K.W. 2007. Contribution of high-frequency information to the acceptance of background noise in listeners with normal and impaired hearing. *Am J Audiol*, 16, 149–156.

Punch J., Joseph A. & Rakerd B. 2004. Most comfortable and uncomfortable loudness levels: Six decades of research. *Am J Audiol*, 13, 144–157.

Recker K.L. & Edwards B.W. 2013. The effect of presentation level on normal-hearing and hearing-impaired listeners' acceptable speech and noise levels. *J Am Acad Audiol*, 24, 17–25.

Rishiq D.A., Harkrider A.W. & Hedrick M.S. 2012. Acceptable Noise Level (ANL) and psychophysical masking. *Am J Audiol*, 21, 199–205.

Rogers D.S., Harkrider A.W., Burchfield S.B. & Nabelek A.K. 2003. The influence of listener's gender on the acceptance of background noise. *J Am Acad Audiol*, 14, 372–382.

Schwartz K.S. & Cox R.M. 2012. Relationship between Acceptable Noise Levels and Hearing Aid Success. Refereed poster presented at the Annual Meeting of the American Auditory Society, Scottsdale, USA, March, 2012.

Tampas J.W. & Harkrider A.W. 2006. Auditory evoked potentials in females with high and low acceptance of background noise when listening to speech. *J Acoust Soc Am*, 119, 1548–1561.

Taylor B. 2008. The Acceptable Noise Level Test as a predictor of real-world hearing aid benefit. *Hear J*, 61(09), 39–42.

von Hapsburg D. & Bahng J. 2006. Acceptance of background noise levels in bilingual (Korean-English) listeners. *J Am Acad Audiol*, 17, 649–658.

Walravens E., Keidser G., Hartley D. & Hickson L. 2012. Does the Acceptable Noise Level predict hearing aid use? UQ-QASA CPD Day, October 13th, 2012, Brisbane.

Comments from Earl Johnson, Melinda Freyaldenhoven Bryan, and Susan Gordon-Hickey

Comments from Earl Johnson

VA Medical Center, Mountain Home, Tennessee, USA

The Discussion Paper brings to the forefront a number of relevant issues related to the acceptance of noise. In addition to questioning whether the acceptable noise level (ANL) procedure has predictive value for determining hearing-aid use, the authors are not convinced that the current methodology for measuring ANL is itself 'acceptable' from either an accuracy or reliability point of view and, consequently, may have less clinical utility than has been previously reported.

Pertinent to the fundamental hypothesis that individuals have an inherent acceptance of noise in the presence of speech, the ANL procedure is perhaps best viewed as originally described in 1991 by Nabelek, Tucker and Letowski. At that time, the procedure was called the tolerated signal-to-noise ratio (SNR), where the wording of 'tolerated' has since been used interchangeably with 'acceptable.'

In the vein of interchangeability, the term 'SNR' was subsequently replaced with 'noise level', where both terms refer to the difference between:

1. The most comfortable level (MCL) for listening to speech; and
2. The maximum background noise level (BNL) while listening to speech in accordance with the instruction set provided by the tolerated SNR (ANL) procedure developers.

From the tolerated SNR viewpoint, the ANL procedure is rather straightforward and less mystic. The concept of a tolerated SNR is most likely readily conceivable and familiar to most all audiologists. From the beginning of our educational and clinical training we, as audiologists, have had speech-recognition-in-noise paradigm training based on the latest evidence. As such, the ANL accomplishment of estimating a tolerated SNR that differs from a separate physical measure of environmental SNR that determines speech recognition performance is logical and, admittedly, clinical appealing. Accordingly, we were well-positioned for reasonable persuasion that ANL might be able to relate well to hearing aid use. And, to be fair with regards to our persuasion, the data was convincingly suggestive of the likely relationship reality between an ANL and hearing aid use (e.g. Nabelek et al, 2006). The timing for such persuasion was excellent because much evidence had in the previous decades already suggested that the SNR loss (impairment) of the patient as well as the speech recognition performance in noise achieved with hearing aids (i.e. objective data) were, somewhat ironically, not related to hearing-aid success (i.e. subjective data) or expected use (Turner & Robb, 1987; Rankovic, 1991; Cox & Alexander, 1991; Bentler et al, 1993a,b; Souza, 2000).

In my opinion, part of the popularity of the ANL procedure may indeed be attributable to its name change from tolerated SNR. Such reasoning is consistent with the perpetual evidence that clever labeling, branding, trademarking, or whatever the 'right wording for the time' is, will always be important to its recognition by others (e.g. Keller, 2008). Tolerated SNR still has much less glam and glitter, if you will, than ANL. Of course, the clinically relevant finding that ANL, a singular measure taking less than two minutes, was reported to account for the classification of hearing-aid use based on reported usage, with 85% accuracy certainly also spurred its popularity (Nabelek et al, 2006). What audiologist, fitting hearing aids one right after other, wouldn't like to have such a simple metric for predicting hearing-aid use? Other simple metrics, establishing self-perception of hearing ability as a predictor of hearing-aid purchase, for example, have also been published (Palmer et al, 2009). After the Nabelek et al (2006) publication, additional research was begun by many researchers around the world to replicate the findings relating ANL to hearing-aid use. Now, with nearly a decade gone by, replication of those original findings have been less than compelling, as was summarized in the current discussion paper.

Beyond just concerns over accuracy of the ANL procedure, the authors of this discussion paper have another point on the distinction between measures of the ANL relating retrospectively versus predicting prospectively hearing-aid use. The authors acknowledge that while some research has retrospectively related hearing-aid usage to measured ANL values (i.e. higher values are associated with less use and vice versa), there have not been any prospective studies of ANL and its predictive validity.

In my limited view of these matters, the ANL procedure can likely gauge an individual's tolerated SNR ratio for a reasonable clinic interpretation. Look at pure-tone thresholds, for example, the known standard deviation is ± 5 dB with the typical Hughson-Westlake procedure. Perhaps acknowledging the variability associated with measuring the ANL is warranted. ANL variability, by definition of its difference calculation between MCL and BNL (two separate measures), is more variable than a singular measure of either MCL or BNL.

Also important to acknowledge is the reality that much of how the ANL procedure gauges tolerated SNR can be influenced by a great number of factors that were discussed throughout this discussion paper. To date, the ANL procedure has been translated reportedly into seven languages. Given that the ANL has been repeatedly reported to be unrelated to an individual's comprehension of the story, one reasonable matter for researchers and clinicians to contemplate is whether it is worthwhile to generate new ANL test material for the near 7000 languages spoken on this planet. There is already the availability of well-reasoned speech-like, but non-semantic signals such as the international speech test signal (ISTS; Holube et al, 2010), and previous demonstration that speech spectrum does not differ across languages and is, in effect, international (ILTASS; Byrne et al, 1994) which can serve as the stimuli for an ANL test.

Ho et al (2013) was the first to preliminarily investigate if the ISTS can serve as the universal stimulus for ANL measures. Moreover, translation and interpretation of the instructions across languages, while possible, has the potential to impact the measured

ANL (e.g. Ho et al, 2013). Therefore, if the predictive relationship of ANL categories (i.e. low, medium, and high) to hearing-aid use, as reported by Nabelek et al (2006), were replicated with confidence and established, for say, American English, then there is not necessarily the potential for application of the same predictive ANL categories to other languages, or possibly even other dialects of the same language.

The next time an article uses the ANL, consider substituting tolerated SNR in its place and then approach the interpretation of the study findings with the scrutiny of scientific approach. If the prediction of hearing-aid use (i.e. a highly behavioral outcome) seems too good to be true, it may well be. From behavioral statistics in psychology, Cohen (1988, 2003) informed researchers and clinicians dealing with behavioral outcomes that accounting for just 25% of the variability (i.e. R-values $\geq .5$) is a notable accomplishment (i.e. high). Even when considering a smorgasbord of 25 predictor variables, Humes (2003) demonstrated that about two-thirds of the variability in hearing-aid outcomes could be accounted for by just a few variables. These variables relating to outcome areas of subjective benefit/satisfaction, speech recognition performance, and hearing-aid usage were namely (1) sound quality, (2) audibility, and (3) hearing-aid experience, respectively. The ANL test was, unfortunately, not included in the Humes (2003) study methodology for predicting hearing-aid usage but, understandably so, as the popularity and expected utility of ANL was not budding until the mid 2000s. It is a curious question to consider whether the ANL may have been able to account for additional variability in hearing-aid outcomes, particularly in the area of hearing-aid usage.

In summary, the ANL procedure should provide a reasonable estimate of the signal-to-noise ratio the listener will tolerate. The estimate should be helpful in discussion with the listener about listening environments and the signal-to-noise ratios he/she may encounter in the real-world. Despite the tolerated SNR of the listener, the primary conventional tasks of any hearing-aid used in treatment are the same. Those primary tasks are to restore effective audibility as a function of frequency and offset SNR loss (i.e. improve speech recognition performance in noise and/or allow the listener to tolerate more adverse SNR listening environments in the context of ANL). Achievement of these tasks should be pursued in ways that maintain subjective benefit and satisfaction of the patient, given pervasive concerns related to issues of a cosmetic nature, occlusion, excessive loudness, etc. that exist when adjusting to hearing aids.

References

Bentler R.A., Niebuhr D.P., Getta J.P. & Anderson C.V. 1993a. Longitudinal study of hearing aid effectiveness I: Objective measures. *J Sp Hear Res*, 36, 808–819.

Bentler R. A., Niebuhr D.P., Getta J.P. & Anderson C.V. 1993b. Longitudinal study of hearing aid effectiveness II: Subjective measures. *J Sp Hear Res*, 36, 820–831.

Byrne D., Dillon H., Tran K., Arlinger S. Wilbraham K. et al. 1994. An international comparison of long-term average speech spectra. *J Acoust Soc Am*, 96, 2108–2120.

Cohen J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. New York, USA: Academic Press.

Cohen J., Cohen P., West S.J. & Aiken L.S. 2003. Applied multiple regression/correlation analysis for the behavioral sciences. 3rd ed. Mahwah, New Jersey: Lawrence Erlbaum.

Cox R.M. & Alexander G.C. 1991. Hearing aid benefit in everyday environments. *Ear Hear*, 12, 127–139.

Ho H., Wu Y., Stangl E., Lentz E. & Bentler R. 2013. The equivalence of acceptable noise level (ANL) with English, Mandarin, and non-semantic speech: A study across the U.S. and Taiwan. *Int J Audiol*, 52, 83–91.

Holube I., Fredelake S., Vlaming M. & Kollmeier B. 2010. Development and analysis of an International Speech Test Signal (ISTS). *Int J Audiol*, 49, 891–903.

Humes L. 2003. Modeling and predicting hearing aid outcomes. Trends in Amplif, 7, 41–75.

Keller K. 2008. *Strategic Brand Management*. 3rd ed. Upper Saddle River, New Jersey: Pearson Prentice-Hall.

Nabelek A.K., Tucker F.M. & Letowski T.R. 1991. Toleration of background noises: Relationship with patterns of hearing aid use by elderly patients. *J Sp Hear Res*, 34, 679–685.

Nabelek A.K., Freyaldenhoven M.C., Tampas J.W., Burchfield S.B. & Muenchen R.A. 2006. Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, 17, 626–639.

Palmer C.V., Solodar H.S., Hurley W.R., Byrne D.C. & Williams K.O. 2009. Self-perception of hearing ability as a strong predictor of hearing aid purchase. *J Am Acad Audiol*, 20, 341–347.

Rankovic C.M. 1991. An application of the articulation index to hearing aid fitting. *J Sp Hear Res*, 34, 391–402.

Souza P.E., Yueh B., Sarubbi M. & Loovis C.F. 2000. Fitting hearing aids with the Articulation Index: Impact on hearing aid effectiveness. *J Rehabil Res Dev*, 37, 473–481.

Turner C.W. & Robb M.P. 1987. Audibility and recognition of stop consonants in normal and hearing impaired subjects. *J Acoust Soc Am*, 81, 1566–1573.

Comments from Melinda Freyaldenhoven Bryan

Louisiana Tech University, Ruston, Louisiana, USA

The Discussion Paper raises some relevant issues regarding ANL. Specifically, it focuses on ANL's differences when using different speech and background noise materials, response formats, and instructions. It further discusses the repeatability and predictive value of ANL. The following are comments, which correspond to each section of the paper.

The ANL procedure

It is very important that clinicians/researchers follow the ANL procedure exactly. Specifically, it is important that clinicians/researchers use the exact instructions for ANL, including the method of bracketing. It is also important that clinicians/researchers obtain, at least two (preferably three) ANL measurements. When completing the repetitions in measurements, it is also important to obtain both MCL and BNL. Then, when determining the patient's ANL, take the median ANL if there is any discrepancy in the three measured ANLs. This should take less than 10 minutes to complete. While some clinicians/researchers may not be completing multiple ANL measurements, the commonplace in audiology is to replicate measurements (i.e. pure tones, SRT, etc.). We rarely settle for one behavioral response to document how a person hears. ANL is no different.

Speech materials

The authors discuss how speech materials (i.e. differences in talker gender, speech rate, and sematic context) might affect ANL. They state that the gender of the talker does not affect ANL, but that speech rate and context of the message might. In the studies that support these findings, the authors state that Helfer (1997) demonstrated that more background noise was accepted when the talker used Clear Speech. It should be noted that Helfer (1997) did not measure ANLs; therefore, this statement is misleading. Helfer (1997) measured listeners' speech perception abilities at various signal-to-noise ratios and found that Clear Speech accompanied with visual cues improves speech recognition abilities in noise. While this strategy might also improve ANLs, this phenomenon has not been investigated. Furthermore, with regard to speech rate, the effect of ANL is inconclusive, and the results from the Brännström et al (2012) study should be viewed with caution in regard to the Swedish text presented. The reason for this is that the BNL instructions when translated were changed significantly. The word 'follow' was changed to 'listen,' changing the meaning of the BNL instructions to 'select the loudness level of the noise that the subjects could accept or 'put up with' without becoming tense or tired while listening to the speech signal for a long time.' The concept of 'listening' is different than 'following' the story, and may have results in lower ANLs for Swedish subjects when compared to other populations. In other words, the difference in the ANL measurements when taken with Swedish and Danish texts might have been due to instructions and not the presentation rate of speech.

Noise materials

In this section of the Discussion Paper, the authors state that 'most studies comparing different noise materials for ANL measurement shows that the noise material has influence on the ANL outcome.' This is not true when looking at the data. In fact, only one study (Brännström et al, 2012) shows that ANL may be affected by the noise material used. The others studies (Nabelek et al, 1991; Crowley & Nabelek 1996; Gordon-Hickey et al, 2012; & Olsen et al, 2012a) found no difference in ANL when using different noise stimuli, while Freyaldenhoven et al (2006) found a statistically significant but not clinical significant difference in ANL for various background noises. The only exception to this is for the inclusive of music as a noise stimulus.

Presentation mode / Response mode

In this section, the authors state that 'direct adjustments by the test subjects should be preferred.' This method or a method using clicks of a button or on a mouse pad should be used to signal the examiner of a requested change. By clicks on a button or mouse pad, I mean one click equals one turn of the audiometer dial. Furthermore, when completing the ANL procedure, do not allow the patient to push the button and hold it to turn the story or the background noise up or down. Instead, always instruct them to click the button or mouse each time they want an increase or decrease. Furthermore, whether using a direct method where the client changes the intensity themselves or using indicator buttons of some kind, these two methods should produce the same results while remaining an inexpensive solution for a clinician/researcher. It should also be noted that the ANL test is no longer available through Cosmos Distributing; it is now sold by Frye Electronics. In my opinion, it is best sold as a CD (and maybe with optional accompanying buttons for signaling), not as a system where the listeners must change the intensity of the story and background noise using dials. The main reason for this is the increase in cost without much increase in the benefit of the test. At least in my geographical area, a decrease in the usage of the hearing in noise test was seen when the cost increased from the purchase of a CD to an entire system.

Instructions

I agree with the authors in that no translation of the instructions should be used. They should be used exactly as they are typed in the Freyaldenhoven et al (2005) or Nabelek et al (2006) manuscripts.

There are two points I would like to make about the instructions. First, it is important to give the listeners verbal instructions as well as written instructions. Second, it is helpful to break the instructions up for patients. For example, when completing ANL, it is helpful to instruct the patient to 'turn the loudness up until it is too loud,' and then allow them to complete this task. Then, instruct them to 'turn the story down until it is too soft.' Allow the patient to complete this; then instruct them to 'select the loudness level that is most comfortable for you.' Do the same three step process with the measurement of BNL. Furthermore, when purchasing the test through Frye Electronics, this is the way the instructions to patients read for clinicians/researcher; however be aware that the instructions are not exactly the same when comparing the instructions published by Frye and the instructions in the Freyaldenhoven et al (2005) and Nabelek et al (2006) manuscripts. Overall, I would recommend using the instructions in the manuscripts with the step by step procedure described by Frye Electronics.

Accuracy and precision of the ANL measurement

In this section of the Discussion Paper, reliability/repeatability of the ANL measurement is discussed. Unfortunately, the concept of reliability in the behavioral sciences is complex and controversial. Whereas some researchers and statisticians recommend completing intraclass correlation coefficients (ICC) and/or analysis of variance statistics for reliability, others recommend the use of standard errors of measurement and/or the coefficient of variation; still others recommend using the coefficient of repeatability (CR). The decision that must be made is one of practically to determine which statistical procedure makes sense for the measurement technique being investigated. Furthermore, because of the dependence on the context in which measurement techniques are used, it can be difficult to make comparisons of reliability/repeatability across different studies.

Furthermore, as stated in the Test Reliability section of a manuscript by Brännström et al (2012; the authors of this Discussion Paper), reliability was determined using intraclass correlation coefficients, which they state, 'indicated high test-retest reliability of MCL, BNL, (and) ANL across repetitions within each condition.' By this statement, I assume the Brännström et al (2012) consider ANL to be both reliable and repeatable. In the present Discussion Paper, however, the authors' state, 'the precision of the ANL measurements is too poor to yield repeatability lower than 3 dB.' The clinical application and interpretation that the authors are placing on the CR statistic is concerning when the literature is inconclusive about the types of statistics that should be completed for reliability/repeatability testing, especially when comparing across studies. Moreover, in order to utilize the CR, the distribution of scores must be normal, a factor that was not addressed in this Discussion Paper for the reported studies. Therefore, the statement that 'the precision of the ANL measurements is poor' is misleading as other statistics have shown the reliability to be high, including research published by the authors of this Discussion Paper. Furthermore, if this type of statement is to be made, it should be backed by a research project which utilizes a combination of approaches to reliability, both ICCs and CRs to be included in the analysis. For more information regarding this issue, please see Olsen et al (2013) and Gordon-Hickey et al (2013), where this issue has been addressed in detail.

The authors also give examples of ranges of differences of various studies (Freyaldenhoven et al, 2006, Olsen et al, 2012b). Because the ANL relies on an individual's response, multiple ANLs (at least two and preferably three) should be completed within a session, and the mean or median ANL used as the listener's ANL. In most listeners, the three ANLs will be very close (within 4 dB of each other); however, there are always exceptions to the rule. Furthermore, to reiterate, the instructions are very important. They must be followed exactly by the testers.

Prediction of future HA use

In this section, the authors state that the 'ANL categories are not very informative.' This is the authors' lack of understanding of the classification of the groups. As stated in the Discussion Paper, successful hearing aid users utilize hearing aids whenever needed whereas unsuccessful users use hearing aids occasionally or not at all. While more research needs to be completed to differentiate the part-time and non-users of hearing aids in the unsuccessful hearing-aid user group, successful hearing aid users utilize hearing aids when they deem they need them and do not use them when they think they don't need them (e.g. reading a book, etc.). Furthermore, other means of categorizing hearing aid users into groups would be based on hours of use, benefit received, and/or satisfaction with hearing aids, all of which have been showed to poor predictors of hearing aid success. Instead of redefining the groups based on different criteria, more research should be completed to differentiate the part-time and non-users of hearing aids that make up the unsuccessful hearing aid group.

Furthermore, Schwartz and Cox (2012) questioned the predictive power of ANL because the predictive power of ANL could not be validated by other standardized hearing-aid outcome measures. This finding is not surprising and does not indicate that the predictive power of ANL is weak. It simply means that ANL measures a different perceptual consequence than the other measurements as there is not another measure of background noise acceptance at this point. Moreover, Hartley et al (2010) found that ANL was not related to hours of hearing-aid use. This finding is consistent with Nabelek et al (2006) findings which acknowledge that patterns of hearing-aid use are more meaningful than hours of hearing-aid use in terms of hearing-aid success. More importantly, hours of hearing-aid use is not a good predictor of hearing-aid success, so the fact that ANL is not related to hours of use is neither surprising nor a bad thing. Also, Walravens et al (2012) supposedly found results in contrast to Nabelek et al (2006) regarding ANLs for successful and unsuccessful hearing-aid users. It is noteworthy to acknowledge that this 'research' appears to be a conference presentation, poster, etc. Furthermore, this 'research' appears not to have been published in a peer reviewed journal, nor has it gone through the peer reviewed process for publication. Before this process is complete, these results should be viewed with extreme skepticism. Lastly, in the Summary of this section, the authors reiterate that 25% of test subjects could be misclassified based on ANL. This is a finding that Dr. Nabelek and colleagues (2007) have acknowledged. Specifically, the listeners with mid-range ANL (7–13 dB) would be the listeners that would be misclassified. That is why more research needs to address the differences in part-time hearing aid users versus full-time and non-users of hearing aids.

A conceptual ANL model

It should be noted that the conceptual ANL model is based on a model that was established based on personal communication. By this, I am assuming the research has been completed but has yet to be presented at a conference or published in a peer-reviewed journal. Personally, I don't think this Discussion Paper is the proper place to introduce a conceptual model for ANL. This model should be introduced in the literature through a peer review process before it is introduced and/or changed in a venue like this. Because this was a personal communication, I cannot review the data that aided in the development of this model and can therefore not make any other comments regarding this issue.

In summary, it should be reiterated that it is extremely important to follow the ANL instructions exactly. Furthermore, the three patterns of hearing-aid use categories need to remain as they were originally designed; however, more research needs to investigate the difference in the three groups, especially the difference in part-time and non-users of hearing aids. As to the question of ANL precision/repeatability, more research should investigate the comparison of reliability statistics for ANL. Moreover, it has been long talked about that one drawback of ANL is the data has always been retrospective in that existing hearing-aid users have been the subjects put into one of three groups. A study investigating the predictive value of ANL before and after hearing aids are fitted would be of great value to the ANL literature.

References

Brännström K.J., Lantz J., Nielsen L.H. & Olsen S.Ø. 2012. Acceptable noise level with Danish, Swedish, and non-semantic speech materials. *Int J Audiol*, 51, 146–156.

Crowley H.J. & Nabelek I.V. 1996. Estimation of client-assessed hearing aid performance based upon unaided variables. *J Speech Hear Res*, 39, 19–27.

Freyaldenhoven M.C., Smiley D.F., Muenchen R.A. & Konrad T.N. 2006. Acceptable noise level: Reliability measures and comparison to preference for background sounds. *J Am Acad Audiol*, 17, 640–648.

Freyaldenhoven M.C., Thelin J.W., Plyler P.N., Nabelek A.K. & Burchfield S.B. 2005. Effect of stimulant medication on the acceptance of background noise in individuals with attention deficit/hyperactivity disorder. *J Am Acad Audiol*, 16, 677–686.

Gordon-Hickey S., Haneghan J.V., Adams E. & Moore R. 2013. Response to Olsen et al (Letter to the Editor). *J Am Acad Audiol*, 24, 241–245.

Gordon-Hickey S., Moore R.E. & Estis J.M. 2012. The impact of listening condition on background noise acceptance for young adults with normal hearing. *J Speech Lang Hear Res*, 55, 1356–1372.

Hartley D., Wanrooy E., Hickson L. & Mitchell P. 2010. Development of an Australian version of the Acceptable Noise Level (ANL) test. Audiology Australia, XIX National Conference, 16–19th of May, 2010, Sydney, Australia.

Helfer K.S. Auditory and auditory-visual perception of clear and conversational speech. 1997. *J Speech Hear Res*, 40, 432–443.

Nabelek A.K., Freyaldenhoven M.C., Tampas J.W., Burchfield S.B. & Muenchen R.A. 2006. Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, 17, 626–639.

Nabelek A.K., Tampas J.W. & Freyaldenhoven M.C. 2007. Further questions about the acceptable noise level test: A response to Dr. Hamill. *J Am Acad Audiol*, 18, 185–187.

Nabelek A.K., Tucker F.M. & Letowski T.R. 1991. Toleration of background noises: Relationship with patterns of hearing-aid use by elderly persons. *J Speech Hear Res*, 34, 679–685.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2012a. The Acceptable Noise Level: Repeatability with Danish and non-semantic speech materials for adults with normal hearing. *Int J Audiol*, 51, 557–563.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2012b. Acceptable Noise Level (ANL) with Danish and non-semantic speech materials in adult hearing aid users. *Int J Audiol*, 51, 678–688.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2013. Intertester reliability of the acceptable noise level (Letter to the Editor). *J Am Acad Audiol*, 24, 241–245.

Schwartz K.S. & Cox R.M. 2012. Relationship between Acceptable Noise Levels and hearing-aid success. Refereed poster presented at the Annual Meeting of the American Auditory Society, Scottsdale, Arizona, March, 2012.

Walravens E., Keidser G., Hartley D. & Hickson L. 2012. Does the Acceptable Noise Level predict hearing aid use? UQ-QASA CPD Day, October 13th, 2012, Brisbane.

Comments from Susan Gordon-Hickey

University of South Alabama, Mobile, Alabama USA

The Discussion Paper raises items for discussion regarding research of Acceptable Noise Levels (ANL). Several of the conclusions coincide with statements made by many previous researchers and are supported by evidence in the literature. However, it seems many of the statements and conclusions made in this Discussion Paper are concerning as they are based on research completed without the use of written and step-by-step (verbal) ANL instructions, are misinterpretations of the ANL literature, are opinion, or are based on work that has not yet been through the process of peer review. A significant portion of this was previously addressed in a Letter to the Editor and Response Letter in the Journal of the American Academy of Audiology (Gordon-Hickey et al, 2013; Olsen et al, 2013). My specific comments regarding the Discussion Paper are below:

The ANL procedure: Number of trials

It is best practice to utilize multiple trials (e.g. three trials) of most comfortable listening level (MCL) and background noise level (BNL) when assessing ANL. I am unable to comment on the statements made in the Discussion Paper regarding Goldman (2009) and Holm & Kastberg (2012) as neither manuscript has been through the process of peer review for publication. I have read the work of Goldman (2009) and believe her findings should be interpreted carefully as participants were instructed to ‘attend’ or ‘not to attend’ to the primary speech during BNL measures. The conclusions of Holm & Kastberg (2012) may be predicated on the use of translated ANL instructions that instruct the listener to ‘listen’ to the story. Neither study appears consistent with ANL instructions which state that the listener should be directed to ‘follow the story.’ ‘Listening’ implies a higher-order auditory function than ‘follow the story.’ These authors may plan to address this when working to publish their work. Although the authors of the Discussion Paper state that the ‘procedure is well established with regard to most of the included steps,’ it is clear that the well established procedures for use of written and step-by-step (verbal) instructions are not universally utilized [see Section 6 ‘Instructions’ as well as the previously mentioned Letter to the Editor and Response Letter in the Journal of the American Academy of Audiology (Gordon-Hickey et al, 2013; Olsen et al, 2013)].

Speech materials

In the Discussion Paper, the authors conclude that ‘Speaker gender does not affect the ANL, but the speech rate and semantic content may have influence on the ANL outcome.’ This conclusion was made after they cite conflicting reports in the literature for all three variables (speaker gender, speech rate, & semantic content). The currently available peer-reviewed publications do not yet allow us to draw definitive conclusions. The authors simply state their opinions in this section. Typically, a straightforward, easy to follow, continuous discourse sample should be employed when measuring ANL (in keeping with the Arizona Travelogue). Unfortunately, it appears that this isn’t always done (e.g. Fredelake et al, 2012).

Noise materials

Based on the current literature, speech babble should be utilized as the background noise for typical ANL measures. Background noise that is more informational in nature (versus energetic) or more fluctuant in nature (versus steady state) appears to affect ANL. Crowley & Nabelek (1996) did not report that ANL differed significantly when measured with speech spectrum noise or speech babble. The Discussion Paper is misleading. Crowley & Nabelek (1996) reported that ANL to speech spectrum noise was 6.9 dB, and that the mean ANL to speech babble was 7.6 dB. No test of differences was completed. In that study, they found that ANL may be helpful in estimating future hearing-aid success. For clarification, conclusions made by myself and my colleagues were that listeners had better (lower) ANLs to multi-talker background conditions compared to the one-talker background conditions. We attribute this to both the informational aspects of the forward one-talker background and the fluctuant nature of the one-talker background. Furthermore, we concluded that the fluctuant nature of a background noise has different effects on ANL than previous reports of fluctuant background noise effects on speech understanding in noise. Since ANL assesses a different phenomenon than speech understanding in noise, this is not surprising.

Presentations mode

It is my understanding that during the initial ANL studies conducted at the University of Tennessee, an investigation of ANLs measured through binaural supraural headphone placement and loudspeaker was conducted with no significant findings. This information should be widely available in peer-reviewed literature. It is largely agreed in the literature that the Nabelek et al (2006) normative data should only be used if traditional ANL measures are completed (i.e. same instructions and presentation mode).

Response mode

ANL is intended to be an easy, straightforward, and clinically feasible pre-hearing-aid fitting measure for use by clinical audiologists. For these reasons, it is important that ANL materials be inexpensive and not need special equipment. With the purchase of a reasonably priced CD, audiologists should be able to utilize this procedure with their hearing-aid patients. Typical audiology clinics are not set up for patients to make direct manipulations of auditory signals presented via the audiometer. For this reason, hand signals are often used and are even recommended by Nabelek et al (2005). If the clinician strictly follows the instructions provided by hand signals from their patient, MCL and BNL measures should accurately represent the patient’s MCL and BNL. Therefore, no special equipment should be needed for measurement of ANL. For this reason, I disagree with the proposal that direct adjustment is needed. However, it should be noted that this determination should be made based on data, not opinion.

Instructions

I agree with the authors that 'with any psycho-physical method, the instructions given are very important for the outcome.' I also agree that 'translation and interpretation of instructions could have a large effect on ANL outcome.' However, not all authors utilize the appropriate combination of written and step-by-step (verbal) instructions (e.g. Brännström et al, 2012; Olsen et al, 2012a,b). Additionally, the translated version of ANL created by the authors of the Discussion Paper changed the final BNL direction; hence, changing the final value to a level that cannot be considered an ANL. I find it concerning that these authors state one opinion in this article and then do the opposite in their own work.

Since ANL measures are psychoacoustic in nature, a fundamental requirement for measuring ANL is use of consistent and accurate instructions (for the audiologist and for the patient). Specifically, there are two forms of instructions provided to patients when ANL measures are made. First, the written instructions are provided to the listener and reviewed in the sound treated room. Next, the audiologist utilizes the complete step-by-step (verbal) instruction set while seated at the audiometer. The step-by-step (verbal) instruction set includes a detailed script for the audiologist to follow as well as instructions regarding step-size changes and methods to respond to patient questions. The Nabelek et al (2006) paper includes instructions that represent the written instructions accurately and provide an overview of the step-by-step (verbal) instructions. Many researchers have erroneously utilized these instructions as both written and step-by-step (verbal) instructions to the listener (e.g. Brännström et al, 2012; Olsen et al, 2012a,b). This is not necessarily a failure on part of these researchers but on the available ANL resources. The written instruction set to provide to the patient (or participant) is often published in the literature (e.g. Rogers et al, 2003 & Freyaldenhoven et al, 2005); however, the specific step-by-step (verbal) instruction set has never been published in an ANL manuscript. The step-by-step (verbal) instruction set was previously available on the University of Tennessee ANL Laboratory website. Additionally, it was published in the CD jacket of the ANL CD when it was available through Cosmos Distributing Incorporated (Nabelek et al, 2005). Unfortunately, Frye Electronics does not appear to provide the original detailed step-by-step (verbal) instruction set. The inconsistent use of written instructions for both written and step-by-step (verbal) instruction to the patient is problematic. For ANL measures to accurately and consistently be measured, the psychoacoustic measurement of MCL and BNL must be precise. For this to occur, all individuals making MCL and BNL measures for assessment of ANL must have access to both instruction sets.

For several reasons, the step-by-step (verbal) instruction set is important for accurate measurement of ANL. First, it provides the audiologist with precise and specific instructions for measuring MCL and BNL. As mentioned, the step-by-step (verbal) instructions provide the audiologist with clear methodological guidelines for set up, calibration, use of written instructions, and use of step-by-step (verbal) instructions. Additionally, the step-by-step (verbal) instruction set includes methods for consistently clarifying confusion that may occur on behalf of the patient. Second, it defines an explicit three-step method for evaluating MCL and BNL. Each step includes a specific instruction as well as methodological instructions. Third, the step-by-step (verbal) instruction set is needed as they guide the audiologist and patient through the bracketing technique used for ANL measures. The use of a bracketing method improves reliability and consistency of MCL and BNL measures (hence, improves reliability of the calculated ANL). The bracketing technique allows the listener to explore a range of speech levels above and below MCL. It also allows the listener to hear background noise levels above and below their level of background noise acceptance (BNL). This process provides the patient with a framework for accurate identification of their MCL and BNL. Last, the step-by-step (verbal) instruction set utilizes very precise language. Deliberate, consistent, and specific instructions are needed because ANL is a psychoacoustic measurement. The specific instruction (3rd step) for evaluation of BNL is imperative to ANL measures. The instruction is 'Now turn the level of the background noise back up the MOST noise that you would be willing to put-up-with and still follow the story for a long period of time without becoming tense or tired.' Slight variations in the wording utilized to instruct a listener may adjust the perception of the listener as to what level must be targeted. Audiologists should not deviate from the language utilized in the step-by-step (verbal) instruction set as this may change the listener's understanding or perception of the intent of the measurements. If the specific instruction is not precisely the same as the instruction set used by Nabelek and colleagues, then we cannot consider the derived measure to be an accurate assessment of BNL (so, ANL calculation would not be accurate).

Because ANL is a psychoacoustic measurement, both written and step-by-step, step-by-step (verbal) instructions must be consistent across researchers and clinicians. Since ANL is a calculation of two measured variables, consistent measurement of each variable is necessary in order to derive an accurate ANL for listeners.

I find it interesting that the authors of the Discussion Paper contend that there is 'no method for precise measurement of ANL'. However, time and time again, researchers utilizing the written and step-by-step (verbal) instructions have concluded that ANL measurements are reliable (e.g. Freyaldenhoven et al, 2006; Gordon-Hickey & Moore, 2007; Nabelek et al 2004, 2006; Nichols & Gordon-Hickey, 2012). The authors of the Discussion Paper have concluded that ANL is reliable in their own work (Brännström et al. 2012)! Again, the overarching problem is that researchers are not all utilizing the same instructions and methods. Once ANL researchers are all operating from the same operational definition, then we move in a positive direction with ANL research.

Factors related to ANL

This is an area of great interest as ANL researchers often work to identify variables that may cause ANL differences across individuals. I am unable to comment on the Walravens et al (2012) study as it has not been published in a peer-reviewed journal.

Accuracy and precision of ANL measurements

The reader is referred to the Letter to the Editor and Response Letter found in the Journal of the American Academy of Audiology for a complete discussion of this topic (Gordon-Hickey et al, 2013; Olsen et al, 2013). Fundamentally, we agree that psychophysical measures can be influenced by the instructions set. I urge the authors to utilize the step-by-step (verbal) instruction set precisely as written by Nabelek et al (2005). This likely plays a significant role in their findings. Additionally, I urge the authors of the Discussion Paper to consider intra-subject variables as a potential for the very few participants who are unable to demonstrate repeatability of measures. Since we do not yet know what specific participant variables influence ANL, the repeatability issues are more likely inherent to the participants and not to the method of measurement (provided that

ANL is measured with written and step-by-step (verbal) instructions utilizing the precise language prescribed by Nabelek).

Prediction of future HA use

The authors of the Discussion Paper state that the ANL categories are 'not very informative'. This is misleading. The mid-range (7–13 dB) ANL group needs continued research so that we can better define this group. However, the high (> 13 dB) and low (< 7 dB) ANL groups are well defined and serve to clearly delineate 'likely to be successful with hearing aids' and 'likely to not be successful with hearing aids' groups. Based on the Nabelek et al (2006) hearing-aid success prediction model, those categorized as having low ANLs have a 75–100% chance at hearing-aid success. Those with high ANLs have a 0–16% chance at hearing-aid success.

The authors of the Discussion Paper argue that the method used to categorize success and non-success with amplification is vague. While I agree that the statement 'I wear my HAs whenever I need them' appears vague, isn't that exactly what we want our patients to do? Don't we consider people who wear their hearing aids when they need them to be successful with amplification? I have a hard time second guessing patients. I am unable to comment on the paper by Schwartz & Cox (2012) as their work is not yet published in a peer-reviewed journal. Last, the authors state that they were unable to replicate (Olsen et al 2012b) the findings of Taylor (2008) who found that ANL 'could predict scores on the International Outcome Inventory for Hearing Aids.' Isn't it possible that a reason the authors were unable to duplicate these finding is due to the specific wording changes made to the instructions provided by Olsen et al (2012b)? In that study, participants were asked to 'listen' to the speech rather than 'follow.' The term 'listen' implies a higher order task than the direction to 'follow.

A conceptual model

At the present time, it seems inappropriate to propose a revised conceptual model when the described model has not yet been published. I would prefer that Wu, Stangl, and Pang have time to publish their work and thoughts prior to another researcher revising their model.

Conclusions

Since ANL is a psychophysical measure, the written and step-by-step (verbal) instructions are fundamental to accurate measurement of MCL and BNL. Use of alternative wording to instruct the patient will result in identification of a level that cannot be considered a BNL (thus, an ANL cannot be calculated). Those using translated versions of ANL must translate the specific terminology of the original directions accurately and completely. Additionally, they must use a primary discourse recording that is similar in nature to the Arizona Travelogue (Nabelek et al 2005). Comparison of ANL findings across studies cannot be made unless the Nabelek et al (2005) procedures were used. The ANL is a reliable and accurate test of background noise acceptance when the Nabelek et al (2005) written and step-by-step (verbal) instructions are followed closely (without ANY deviation).

References

Brännström K.J., Lantz J., Nielsen L.H. & Olsen S.Ø . 2012. Acceptable noise level with Danish, Swedish, and non-semantic speech materials. *Int J Audiol*, 51, 146–156.

Crowley H.J. & Nabelek I.V. 1996. Estimation of client-assessed hearing aid performance based upon unaided variables. *J Speech Hear Res*, 39, 19–27.

Fredelake S., Holube I., Schlueter A. & Hansen M. 2012. Measurement and prediction of the acceptable noise level for single-microphone noise reduction algorithms. *Int J Audiol*, 51, 299–308.

Freyaldenhoven M.C., Smiley D.F., Muenchen R.A. & Konrad T.N. 2006. Acceptable noise level: Reliability measures and comparison to preference for background sounds. *J Am Acad Audiol*, 17, 640–648.

Freyaldenhoven M.C., Thelin J.W., Plyler P.N., Nabelek A.K. & Burchfield S.B. 2005. Effect of stimulant medication on the acceptance of background noise in individuals with attention deficit/hyperactivity disorder. *J Am Acad Audiol*, 16, 677–686.

Goldman J.J. 2009. The effects of test method, alternate types of target stimuli and attention on Acceptable Noise Level (ANL) scores in normal-hearing listeners. Doctoral dissertation, Communication Sciences and Disorders, James Madison University.

Gordon-Hickey S., Haneghan J.V., Adams E. & Moore, R. 2013. Response to Olsen et al (Letter to the Editor). *J Am Acad Audiol*, 24, 241–245.

Gordon-Hickey S. & Moore R. 2007. Influence of music and music preference on acceptable noise levels. *J Am Acad Audiol*, 18, 417–428.

Holm L. & Kastberg T. 2012. Stabilitet för Acceptable Noise Level (ANL) hos normalhörande vuxna personer vid upprepade matningar inom samma testsession och dess relation till arbetsminneskapacitet. Vetenskapligt arbete, Avdelningen for logopedi, foniatri och audiologi Institutionen för kliniska vetenskaper, Medicinska Fakulteten, Lunds Universitet, Lund [in Swedish].

Nabelek A.K., Freyaldenhoven M.C., Tampas J.W. & Burchfield S.B. 2005. Quality Recordings for the Hearing Health Care Industry: Acceptable Noise Level Test (ANL). Compact Disc. Cosmos Distributing Incorporated.

Nabelek A.K., Freyaldenhoven M.C., Tampas J.W., Burchfield S.B., Muenchen R.A. 2006. Acceptable noise level as a predictor of hearing-aid use. *J Am Acad Audiol*, 17, 626–639.

Nabelek A.K., Tampas J.W. & Burchfield S.B. 2004. Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. *J Speech Lang Hear Res*, 47, 1001–1011.

Nichols A.C. & Gordon-Hickey S. 2012. The relationship of locus of control, self-control, and acceptable noise levels for young listeners with normal hearing. *Int J Audiol*, 51, 353–359.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2012a. The Acceptable Noise Level: Repeatability with Danish and non-semantic speech materials for adults with normal hearing. *Int J Audiol*, 51, 557–563.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2012b. Acceptable Noise Level (ANL) with Danish and non-semantic speech materials in adult hearing aid users. *Int J Audiol*, 51, 678–688.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2013. Intertester reliability of the acceptable noise level (Letter to the Editor). *J Am Acad Audiol*, 24, 241–245.

Rogers D.S., Harkrider A.W., Burchfield S.B. & Nabelek A.K. 2003. The influence of listener's gender on the acceptance of background noise. *J Am Acad Audiol*, 14, 372–382.

Schwartz K.S. & Cox R.M. 2012. Relationship between Acceptable Noise Levels and Hearing Aid Success. Refereed poster presented at the Annual Meeting of the American Auditory Society, Scottsdale, Arizona, March, 2012.

Taylor B. 2008. The Acceptable Noise Level Test as a predictor of real-world hearing aid benefit. Hear J, 61(09), 39–42.

Walravens E., Keidser G., Hartley D. & Hickson L. 2012. Does the acceptable noise level predict hearing-aid use? UQ-QASA CPD Day, October 13th, 2012, Brisbane.

Author's responses

We would like to thank Drs. Johnson, Freyaldenhoven Bryan, and Gordon-Hickey very much for their willingness to comment on our ANL discussion paper. We find it valuable to have a good and open academic discussion of ANL and its potential use, as well as methods to improve the measure. In the following we will address the issues raised by the three discussants ordered by topic.

ANL expressed as a SRT

Dr. Johnson's thoughts about the naming of ANL are very interesting, and we agree that the interpretation of the outcome expressed as a SNR would be straightforward as opposed to the ANL. ANL expressed as dB SNR would make it easier to understand why a high ANL indicates something negative (the listener will not become a full-time user), while a low ANL indicates something positive (the listener will become a full-time user).

The ANL procedures and instructions

As stressed by Drs. Freyaldenhoven Bryan and Gordon-Hickey, ANL should be obtained using a well-defined methodology, but even when doing so, it is not certain that the existing method yields precise results. For example we are not convinced that two or three measurements may be sufficient to obtain precise results. Parts of the dissertation by Holm & Kastberg (2012) have recently been accepted for publication (Brännström et al, 2013). The results show that the CR continues to decrease through four ANL measurements based on three repetitions each. Based on that research it seems that ideally three ANL test runs consisting of three measurements should be done before the ANL is finally calculated on the fourth ANL run. Therefore the mean of two or three ANL measurements may not be sufficient. However, this remains to be studied using the American version of the ANL test among others.

In contrast to Dr Freyaldenhoven Bryan, we think that the ANL instruction set can and should be translated. Listeners should have a very exact instruction on how the ANL should be carried out and that instruction must be in the listener's mother tongue. There is no way around this if the ANL should be used internationally, but as we stressed in our discussion paper, great care has to be taken when translating the ANL instructions. We do not agree with Dr. Gordon-Hickey that our translation have changed the final BNL direction. The Danish instructions are an exact translation of Nabelek et al's (2004) instruction set. The Swedish instructions have only a small and probably non-significant deviation, as we discuss in the 'Prediction of Future HA Use' section of this response.

Factors that affect the ANL

We agree with Drs. Freyaldenhoven Bryan and Gordon-Hickey that definitive conclusions cannot be drawn about the effects of speaker gender, speech rate, and semantic content on the ANL from the available peer-reviewed literature. However, in single studies using the speech and noise signals with continuous discourse and a multispeaker noise as test signals, the possible variability introduced by speaker gender, speech rate, and semantic content should not influence the association between ANL and hearing-aid use: Any changes incurred by these factors can potentially alter the actual ANL values in the distribution but probably not the intersubject variability. Thus, for groups of subjects, these factors should not affect any relative relationship between the ANL and hearing-aid use. A study using a consistent set-up, consistent instructions, and sufficient sample size should be able to reveal any relationship between ANL and hearing-aid use.

According to Dr. Freyaldenhoven Bryan, the use of the word 'listen' in one instruction set and 'follow' in the other was the reason for the differences between ANLs found in Denmark and Sweden (Brännström et al, 2012). As discussed further in the 'Prediction of Future HA Use' section, we don't see this difference as the explanation.

As suggested by Dr. Johnson, it might be a good idea to use the non-semantic ISTS speech signal internationally to allow for comparison across language boarders. As background noise we would propose to use the ANL babble, since it is the most frequently used noise type. Other types might be used for specific purposes.

Some of the ANL variability might also, as suggested by Dr. Johnson, be owing to the very nature of ANL being based on the two separate measures: MCL and BNL. This needs further examination.

Presentation and response modes

We are convinced that some of the response modes used in the past may have introduced bias to the measurements. A good way of avoiding some of the bias is to exclude the tester from the link between test person and equipment. That can be done when the test person directly adjusts the level of speech and noise. The method has to be as intuitive as possible, and we do not think that asking the test person to use mouse clicks for the bracketing of the ANL as suggested by Dr. Freyaldenhoven Bryan is intuitive. Using an intuitive method with limited bias from the tester has in our opinion priority above the cost of the method.

Accuracy and precision of the ANL measurement

When we started our ANL research we had a simplistic view of the robustness of the ANL measurement therefore we based our instructions on the information provided by Nabelek et al (2004). We have now come to the understanding that the criterion validity of the ANL must be unclear and that these issues may be crucial to the ANL result. Therefore, the clarification of the instructions made by Dr. Gordon-Hickey is highly appreciated.

As Dr. Gordon-Hickey states, there are currently no studies available that examine the effects of omitting the step-by-step verbal instructions on the ANL. We agree that there will probably be an effect, but future studies need to answer this question. It is possible that this discrepancy could be the reason for our study not showing a relation between ANL and hearing-aid use (Olsen et al, 2012). However, a more plausible explanation is the poor repeatability seen for the ANL test, e.g the large CRs that are encountered for both the Danish and the American version (Olsen et al, 2012; Gordon-Hickey et al, 2012a; Olsen et al, 2013). Also as noted by Dr. Gordon-Hickey, it is our opinion that the examiner is a potential source of bias in the ANL in the same manner as it is for pure-tone audiometry, but we strongly agree that this needs to be tested before any solid conclusions can be drawn.

It is correct as Drs. Gordon-Hickey and Freyaldenhoven Bryan write, that our conclusion was that the ANL was indeed reliable (Brännström et al, 2012) and this conclusion was made from the best of our understanding then. However, our later research has led us to question the use of simple correlations, repeated measures ANOVAs, or intraclass correlation to analyse the reliability of a method. Simple correlations (e.g Pearson's correlation coefficient) used by some researchers (e.g. Nabelek et al, 2004; Gordon-Hickey & Moore, 2007; Gordon-Hickey et al, 2012a) to draw this conclusion is in our opinion not an appropriate statistical method (Bland & Altman, 1986). For example two measures (x and y) will demonstrate a high correlation coefficient if the relationship between the measures is y = 0.5x, but in terms of agreement between the two measures they will only agree if the relation is y = x (note that the same would also be true if we used a constant offset/bias where y = 0.5x + b). Therefore, the use of correlation coefficients is not an appropriate method when examining the reliability of repeated measures.

ICC may show whether ANLs obtained at two sessions by one or more testers are associated, but it does not show the precision of the ANL. Weir (2005) states, that if between-subject variability is high the ICC may be high even when test retest differences are large (the test has a low precision). If on the other hand the test retest differences are small (the test has a good precision) the ICC may be low if the between-subject variability is small. As an example please refer to Figure 2C in the study by Gordon-Hickey et al (2012a), in which the testers followed the instruction set faithfully. The figure is a scatter plot of the ANL results obtained by testers B and C. Since ICC was significant and one-way ANOVAs were not, the authors concluded that ANLs are reliable across testers. In contrast to this, Figure 2C (Gordon-Hickey et al, 2012a) reveals large individual differences between the ANLs obtained by the testers B and C. The reason why the ICC reported by Gordon-Hickey et al (2012a) is statistically significant despite the large individual differences between measurements might be the large between-subjects variation (Weir, 2005). Information about the size of the individual differences can easily be achieved by calculating the CR. ICC and CR show different aspects of repeated measures and therefore we do not see ICC as a golden standard to which the CR should be compared as suggested by Dr. Freyaldenhoven Bryan. However, we do agree that it would be good practice to provide both CR and ICC in future research reports.

In the case of repeated measures ANOVA, as used by some previous studies as a measure of reliability (e.g. Freyaldenhoven et al, 2006; Brännström et al, 2012; Gordon-Hickey et al, 2012a), the differences between two repeated measures may be very small and not significant, but this analysis does not provide any information on the performance of individual subjects. Since the mean is used in this analysis, a single subject may actually have a large ANL in the first measurement and a very low in the second but it may not significantly affect the mean difference between the measurements. The CR is more attractive since it can be used to relate the performance of a single subject with the performance of the group and what would be a significant difference from one measurement to another. Hence, the CR is directly related to the performance and demonstrates what is a relevant difference in the unit measured (dB).

We do not share Dr. Freyaldenhoven Bryans opinion that it is misleading to state that 'the precision of the ANL measurements is poor'. As we have shown, the CRs of the repeated measurements in the study by Gordon-Hickey et al (2012a) were up to 7.6 dB (Olsen et al, 2013). We agree with Gordon-Hickey & Moore (2008) who stated, that 1–3 dB difference in ANL might influence the prediction of hearing aid outcome, and that is why the ANL having a CR of 7.6 has too poor precision to be used for prediction of hearing-aid use.

The explanation for the intra-subject variability seen in the ANL literature may lie in both procedures and intra-subject variables but future studies are required to examine this. However, so far, and despite a large number of studies, few intra-subject variables have been indentified and this may depend on poor repeatability.

Prediction of future HA use

We agree with Dr. Gordon-Hickey that the comparison between the normative data reported by Nabelek et al (2006) and data obtained using different versions of the ANL should be avoided due to the many issues we have put forward in our discussion paper. However, if the ANL is an inherent characteristic of the individual as suggested by many (Nabelek et al, 1991; Rogers et al, 2003; Nabelek et al, 2006; Tampas & Harkrider, 2006) the presentation mode should not influence any possible relation with hearing aid use pattern.

We do not agree with Dr. Freyaldenhoven Bryan that we lack understanding of the classification of the ANL groups. Dr. Gordon-Hickey shares our opinion that the statement 'I wear my HAs whenever I need them' appears vague. It is our belief that our patients may think that they use their hearing aids whenever they need them. In many cases, this could be the same thing as that they use them when they would benefit from using them but most likely not in all cases. It is possible to think that you use your hearing aids when needed and not being what we define as a successful hearing-aid user using e.g. the International Outcome Inventory for Hearing Aids (Cox & Alexander, 2002; Cox et al, 2002).

It is possible as mentioned by Dr Gordon-Hickey that the wording of our instructions have influenced our ability to replicate the findings of Taylor (2008). However, it is an unanswered question whether the meaning of 'follow the story for a long time' is different from 'listening to the story for a long time' in Danish. In the widely used instructions by Nabelek et al (2004) the word listen is used four times ('You will listen to a story…', '….listening to a radio.' 'You will listen to the same story…', 'After you have listened…') and the word follow is used twice ('…while following the story.'). In our opinion, in this context there is no difference between the two words and both tasks (to listen and to follow) require higher order processes which we refer to as listening rather than hearing.

It is very positive that both Drs. Freyaldenhoven Bryan and Johnson support that the ANL predicting model should be tested on a population other than the population on which the model was created. As stated by Dr. Johnson, it is also important to keep in mind that even if the ANL predictive model proves to work with American English, then this is not necessarily the case with other language versions.

A conceptual ANL model

The model presented in our discussion paper is a tool for discussing ANL. Our model was based on the one proposed by Wu, Stangl, & Pang (personal communication, 2013), which has now been accepted for publication (Wu et al, 2013). We do not think that our model holds the final truth about ANL, but we have included as much knowledge about ANL as possible. We think of our model as a stepping stone in the stream of ANL information, in which Wu et al's (2013) model was the first stepping stone. We hope that other researches will follow our example and remove some of the stepping stones while adding others, hence giving a better understanding of ANL while trying to improve the ANL method.

ANL and the future

As Dr. Johnson summarizes, the ANL should help the audiologist when discussing listening environments and real-world SNRs with the patient. We hope that the debate raised by our discussion paper will continue and lead to more research aimed at improving the ANL concept, so it can fulfill this achievement. If the ANL shall be of continued use for clinicians, it will need to be developed into a precise measure that can be proven to accurately point out to future hearing-aid users the problems that will prevent them from being satisfied hearing-aid users. If this cannot be achieved, little will in our opinion be gained by measuring ANL.

References

Bland J.M. & Altman D.G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307–310.

Brännström K.J., Holm L., Kastberg T. & Olsen S.Ø. 2013. The acceptable noise level: The effect of repeated measurements. *Int J Audiol*, In press.

Brännström K.J., Lantz J., Holme Nielsen L. & Olsen S.O. 2012. Acceptable Noise Level with Danish, Swedish, and non-semantic speech materials. *Int J Audiol*, 51, 146–156.

Cox R.M. & Alexander G.C. 2002. The International Outcome Inventory for Hearing Aids (IOI-HA): Psychometric properties of the English version. *Int J Audiol*, 41, 30–35.

Cox R.M., Stephens D. & Kramer S.E. 2002. Translations of the International Outcome inventory for Hearing Aids (IOI-HA). *Int J Audiol*, 41, 3–26.

Freyaldenhoven M.C., Smiley D.F., Muenchen R.A. & Konrad T.N. 2006. Acceptable Noise Level: Reliability measures and comparison to preference for background sounds. *J Am Acad Audiol*, 17, 640–648.

Gordon-Hickey S., Adams E., Moore R., Gaal A., Berry K. et al. 2012a. Intertester reliability of the Acceptable Noise Level. *J Am Acad Audiol*, 23, 534–541.

Gordon-Hickey S. & Moore R.E. 2007. Influence of music and music preference on acceptable noise levels in listeners with normal hearing. *J Am Acad Audiol*, 18, 417–427.

Gordon-Hickey S., Moore R.E. & Estis J.M. 2012b. The impact of listening condition on background noise acceptance for young adults with normal hearing. *J Speech Lang Hear Res*, 55, 1356–1372.

Gordon-Hickey S. & Moore R.E. 2008. Acceptance of noise with intelligible, reversed, and unfamiliar primary discourse. *Am J Audiol*,17, 129–135.

Holm L. & Kastberg T. 2012. Stabilitet för Acceptable Noise Level (ANL) hos normalhörande vuxna personer vid upprepade mätningar inom samma testsession och dess relation till arbetsminneskapacitet. Vetenskapligt arbete, Avdelningen för logopedi, foniatri och audiologi Institutionen för kliniska vetenskaper, Medicinska Fakulteten, Lunds Universitet, Lund. [In Swedish]

Nabelek A.K., Freyaldenhoven M.C., Tampas J.W., Burchfield S.B. & Muenchen R.A. 2006. Acceptable noise level as a predictor of hearing aid use. *J Am Acad Audiol*, 17, 626–639.

Nabelek A.K., Tampas J.W. & Burchfield S.B. 2004. Comparison of speech perception in background noise with acceptance of background noise in aided and unaided conditions. *J Speech Lang Hear Res*, 47, 1001–1011.

Nabelek A.K., Tucker F.M. & Letowski T.R. 1991. Toleration of background noises: Relationship with patterns of hearing aid use by elderly persons. *J Speech Hear Res*, 34, 679–685.

Rogers D.S., Harkrider A.W., Burchfield S.B. & Nabelek A.K. 2003. The influence of listener's gender on the acceptance of background noise. *J Am Acad Audiol*, 14, 372–382.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2012. Acceptable Noise Level (ANL) with Danish and non-semantic speech materials in adult hearing aid users. *Int J Audiol*, 51,678–688.

Olsen S.Ø., Nielsen L.H., Lantz J. & Brännström K.J. 2013. Intertester reliability of the acceptable noise level. *J Am Acad Audiol*, 24, 241–243.

Tampas J.W. & Harkrider A.W. 2006. Auditory evoked potentials in females with high and low acceptance of background noise when listening to speech. *J Acoust Soc Am*, 119, 1548–1561.

Taylor B. 2008. The Acceptable Noise Level Test as a predictor of real-world hearing aid benefit. *Hear J*, 61, 39–42.

Weir J.P. 2005. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*, 19, 231–240.

Wu Y., Stangl E., Pang C. & Zhang X. 2013. The effect of audiovisual and binaural listening on the acceptable noise level (ANL): Establishing an ANL conceptual model. *J Am Acad Audiol*. In press.